



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

텍스트마이닝을 이용한 메르스 관련
정부 보도 자료에 따른
언론 및 공중의 반응 분석

연세대학교 보건대학원
보건통계학과 보건정보관리전공
정 진 기

텍스트마이닝을 이용한 메르스 관련
정부 보도 자료에 따른
언론 및 공중의 반응 분석




지도 남 정 모 교수

이 논문을 보건학 석사 학위 논문으로 제출함

2016년 12월 일

연세대학교 보건대학원
보건통계학과 보건정보관리전공
정 진 기

정진기의 보건학 석사 학위논문을 인준함

심사위원 남 경 모 
심사위원 박 소 희 
심사위원 김 희 진 

연세대학교 보건대학원

2016년 12월 일

감사의 말씀

보건대학원 보건통계학과 보건보관리전공으로 입학하기까지 많은 우여 곡절이 있었습니다. 세 번의 도전 끝에 유일한 지원자로 보건대학원의 가족이 되었고 열정과 자부심을 가지고 학업에 임하였습니다. 훌륭한 교수님들과 선배님들의 뒤를 이어 보건대학원의 동창이 되었다는 사실이 매우 뜻 깊고 자랑스럽습니다.

석사과정을 무사히 끝마칠 수 있도록 이끌어 주신 남정모 교수님, 항상 귀한 가르침을 주신 박소희 교수님, 그리고 세심하게 실질적인 조언을 해주신 김희진 교수님께 머리 숙여 깊은 감사의 말씀드립니다.

일과 학업을 병행할 수 있도록 배려해 주시고, 늘 관심과 애정으로 지켜봐 주신 의료원장님과 의료원 가족들에게도 감사의 마음 전합니다. 2년 6개월 동안 든든히 옆에서 도와준 동기 선희누나, 순동이형, 은지, 은비와 강정모 선배님, 차윤희 선배님, 이정호 선배님께도 감사드립니다.

세상에서 가장 존경하는 아버지, 늘 기도해주시는 어머니, 항상 큰 힘이 되는 동생 진이에게 감사와 사랑의 마음을 전합니다. 논문이 완성되기까지 옆에서 큰 도움을 준 아내 자유희와 장인어른, 장모님께도 큰 감사와 사랑의 마음을 전합니다.

2016년 12월

정 진 기 올림

차 례

국문요약	v
I. 서 론	1
1. 연구배경 및 필요성	1
2. 연구 목적	3
II. 이론적 배경	4
1. 메르스 특성과 환자발생 양상	4
2. 텍스트 마이닝 분석의 개념	7
3. 관련 연구	11
III. 연구방법	15
1. 연구방법의 설정	15
2. 연구 모형	16
3. 분석 방법	18
IV. 연구결과	21
1. 분석자료의 일반적 특성	21
2. 텍스트 마이닝 결과	26
3. 코사인 유사도 분석	40
V. 고찰 및 결론	48

참고문헌	53
Abstract	58

표 차 례

표 1. 연구대상 자료의 일반적 특성	22
표 2. 단어 빈도(Term Frequency)	26
표 3. 초기 대응단계 단어 빈도(Term Frequency)	27
표 4. 적극 대응단계 단어 빈도(Term Frequency)	29
표 5. 후기 대응단계 단어 빈도(Term Frequency)	30
표 6 중복 단어 결과	41
표 7 단계별, 그룹별 중복 단어 벤 다이어그램	44
표 8 코사인 유사도 분석 결과	47

그림 차례

그림 1. 연구 모형.	17
그림 2. 단계별 메르스 관련 정부 보도자료 건수.	23
그림 3. 단계별 메르스 관련 언론 보도자료 건수.	24
그림 4. 단계별 메르스 관련 언론 보도자료 댓글 건수.	25
그림 5. 초기 대응단계 정부 보도자료 워드클라우드(WordCloud).	32
그림 6. 초기 대응단계 언론 보도자료 워드클라우드(WordCloud).	33
그림 7. 초기 대응단계 공중의 반응 워드클라우드(WordCloud).	34
그림 8. 적극 대응단계 정부 보도자료 워드클라우드(WordCloud).	35
그림 9. 적극 대응단계 언론 보도자료 워드클라우드(WordCloud).	36
그림 10. 적극 대응단계 공중의 반응 워드클라우드(WordCloud).	37
그림 11. 후기 대응단계 정부 보도자료 워드클라우드(WordCloud).	38
그림 12. 후기 대응단계 언론 보도자료 워드클라우드(WordCloud).	39
그림 13. 후기 대응단계 공중의 반응 워드클라우드(WordCloud).	40

국 문 요 약

텍스트마이닝을 이용한 메르스 관련 정부 보도 자료에 따른 언론 및 공중의 반응 분석

연구 배경

최근 몇 년간 전 세계는 중증급성호흡기증후군(Severe Acute Respiratory, SARS), 에볼라(Ebola), 중동호흡기증후군(Middle East Respiratory Syndrome, MERS), 지카바이러스(Zika virus) 등 감염병으로 인한 위기 대응의 중요성이 부각되고 있다. 2015년 5월에 발생한 메르스 사태는 정부의 위기 커뮤니케이션이 실패한 대표 사례로 회자되고 있다. 정부는 메르스 위험의 진상과 확산의 경로를 충분히 국민들에게 설명하지 않았고 감염자가 발생한 병원의 이름을 공개하는 것을 꺼렸다. 이에 시민들은 자발적으로 사회연결망서비스(Social Network Service, SNS)를 통하여 메르스에 대한 위험 정보를 자발적으로 교환하기 시작했다. 그러자 정부는 메르스에 대한 부정확한 정보를 확산시키는 사람들을 엄단하겠다고 선언하였다. 하지만 이러한 정부의 발표는 오히려 국민들의 공분을 자아내게 했다(김은성, 2015). 본 연구는 메르스 관련 정부의 대응 메시지를 분석하여 정부가 위기를 어떻게 인식하였고 대응하였는지, 그에 따른 언론 및 공중의 반응은 어떠했는지 살펴보고자 하였다.

대상 및 방법

연구 대상자료는 정부 공식 보고에 따라 메르스 확산 양상에 따라 초기 대응단계(2015년 5월 20일~6월 8일), 적극 대응단계(2016년 6월 9일~7월 27일),

후기 대응 및 복구단계(2015년 7월 28일~12월 23일)로 시기를 나누고, 중앙정부 공식 홈페이지에 게재된 메르스 관련 보도자료 총 286건, 언론 보도자료 137건, 공중의 반응(언론기사 댓글) 411건을 수집하였다. 이를 텍스트 마이닝을 통해 전처리 및 분석하였다. 단어 빈도(Term Frequency), 워드 클라우드(Word Cloud), 코사인 유사도(Cosine Similarity) 분석을 통해 정부 보도자료, 언론 보도자료, 공중의 반응의 양상과 유사도를 분석하였다.

연구결과

메르스 사태의 주요 시기별로 단어빈도 및 워드클라우드 분석 결과 정부는 메르스 현황 파악, 병원명 공개, 대책 마련 및 복구대응에 집중하였다. 언론은 메르스 현황 및 정부의 대응 및 종식에 주목하였다. 공중은 주로 정부의 대응에 비판하며 메르스 현황, 정부의 대응, 희망적 메시지, 반성 등이 나타났다. 빈도를 제외한 단어 중복율 분석 결과 초기 대응단계, 적극 대응단계, 후기 대응 및 복구단계 모두 단계에서 공중-정부, 언론-공중, 정부-언론 순으로 높게 나타났음을 확인하였다. 정부와 언론의 단어 중복율이 가장 낮았으며 이는 정부가 언론이 원하는 정보를 적절하게 제공하지 않았다고 볼 수 있다. 코사인 유사도 분석 결과 초기 대응단계, 적극 대응단계, 후기 대응 및 복구단계 모두 정부 보도자료와 언론 보도자료가 가장 높은 코사인 유사도를 나타냈으며, 정부 보도자료와 공중의 반응의 유사도가 가장 낮은 유사도를 나타냈다. 즉, 정부가 전달한 메시지와 공중이 원하는 메시지의 유사도가 비교적 낮았다.

결론

메르스 확산 시기에 따라 정부 보도자료에 따른 언론 및 공중의 반응을 살펴본 결과 정부의 위기 대응 메시지와 공중의 반응에 대한 유사도가 낮은 것

으로 나타났다. 이는 공중이 원하는 정보를 정부에서 적절히 대응하지 못한 것이라고 볼 수 있다. 메르스 초기 단계에 정부는 정보를 제한적으로 공개하였으며 공중이 원하는 정보에 집중하지 않았고 일방적인 현황보고 중심으로 대응이 이루어졌다고 볼 수 있다. 연구 대상 자료 수집에 대한 한계점이 있지만 메르스라는 국가적 재난 상황에서 정부의 대응 메시지와 그에 따른 언론 및 공중의 반응이 통계적으로 차이가 있으며, 향후 정부의 위기 대응 메시지가 언론 및 공중이 원하는 정보를 적절히 제공해야한다는 결과를 도출했다는 데에 의의가 있다.

I. 서 론

1. 연구배경 및 필요성

지난 2015년 5월 20일 첫 메르스(Middle East Respiratory Syndrome, MERS) 환자가 확진되면서 우리나라에 메르스 유행이 시작되었다. 2015년 12월 23일 메르스 유행 종료를 선언한 날까지 환자 186명, 사망자 38명이 발생하고 16,693명을 격리하였다(메르스 백서, 2015).

“감염병의 예방 및 관리에 관한 법률” 제6조는 “국민의 권리와 책무”라는 표제로 “① 국민은 국가와 지방자치단체의 감염병 예방 및 관리를 위한 활동에 적극 협조하여야 한다.”, “② 국민은 감염병 발생 상황, 감염병 예방 및 관리 등에 관한 정보와 대응방법을 알 권리가 있다.” 라고 규정하고 있다. 정부는 메르스 사태 초기에 세계보건기구(World Health Organization, WHO) 자료만을 바탕으로 메르스의 감염력이 낮고 독성은 감기수준이라고 발표하였다(김은성, 2015). 정부는 첫 확진 환자 발생 10일 만에 메르스 핫라인(109)을 개설하고 위기소통을 시작하였으나 국민들의 과도한 불안이나 오해를 막기 위해 메르스 환자가 경유한 병원명을 비공개하는 방침을 고수하였다(연합뉴스, 2015). 국민들의 혼란과 불안은 SNS 등을 통해 퍼지는 괴담 등으로 인하여 점점 증가하였다. 결국 첫 확진 환자 발생 2주 만에 메르스 환자 경유 병원이 공개되었지만 한국-세계보건기구 합동평가단은 한국 정부가 정보 공개를 늦춘 탓에 초기 메르스 방역 정책의 실패를 불러왔다고 평가했다(연합뉴스, 2015). 확진환자 발생·경유 의료기관명 공개와 같이 국민이 궁금해 하는 사항에 대해 적시에 대응을 하지 못해 국민의 불신은 높아져 갔으며, 기존의 위기소통 방식으로 대응하는 데 한계가 노출되었다(메르스 백서, 2015).

이처럼 메르스 관련 정부의 위기소통에 대해 많은 비판과 문제가 제기되었고 이와 관련된 다양한 선행연구가 수행되었다. 뿐만 아니라 메르스 관련 매스미디어의 언론보도를 분석하는 연구와 트위터 등의 SNS 및 기사 댓글 등의 메시지를 대상으로 공중의 반응에 대한 선행연구도 진행되었다.

정부가 국민에게 메르스 상황, 예방 및 관리 등에 관한 정보와 대응방법을 적절하게 제공했는지 여부를 판단하기 위해 정부의 메시지가 언론과 공중의 메시지와 얼마나 일치했는지 알아보는 것이 필요하지만 정부-언론-공중의 반응을 통합적으로 비교 및 분석한 연구는 진행되지 않았다.

이에 본 연구에서는 주요 시기별로 정부-언론-공중이 주로 사용했던 메시지는 무엇인지, 정부-언론-공중의 메시지가 얼마나 일치했는지 알아보고 이를 통해 정부는 언론과 공중이 원하는 정보를 얼마나 적절하게 제공했는지 알아보고자 한다.

2. 연구목적

본 연구의 목적은 메르스 기간 동안 정부의 보도 자료가 어떠한 정보와 대응방법을 제시하였는지, 그에 따른 언론 및 공중의 반응은 어떠한지 알아보고 이를 통해 정부는 언론과 공중이 원하는 정보를 적절히 제공하였는지 확인하는 것이다.

세부적인 연구 목적은 다음과 같다.

첫째, 메르스 유행기간 동안 주요 시기별로 정부 보도자료, 언론 보도자료, 공중의 반응을 텍스트 마이닝을 통해 분석한다. 이를 통해 정부, 언론, 공중이 메르스 주요 시기별로 주요 사용 단어의 양상을 분석한다.

둘째, 텍스트 마이닝 결과를 바탕으로 정부, 언론, 공중이 사용한 단어의 유사도를 분석한다. 정부, 언론, 공중이 주요 시기별로 메르스에 대한 관심사가 얼마나 유사했는지 통계 분석을 통해 알아본다.

II. 이론적 배경

1. 메르스 특성과 환자발생 양상

1) 메르스 바이러스의 특성

메르스는 메르스 코로나바이러스(Middle East Respiratory Syndrome Coronavirus, MERS-CoV)에 의한 호흡기 감염증이다. 메르스 코로나바이러스는 Zaki 등(2012)이 중증 폐렴과 신부전으로 사망한 60대 남성으로부터 최초로 발견하였다. 2013년 5월 국제바이러스 분류위원회(International Committee on Taxonomy of Viruses, ICTV)에서는 이 신종 코로나바이러스를 메르스 코로나바이러스라 명명하였다(질병관리본부, 2015).

메르스 코로나바이러스에 감염되면 환자의 대부분은 중증급성하기도질환(폐렴) 증세를 보이고, 일부는 무증상을 나타내거나 경한 급성상기도질환을 나타낸다. 발열, 기침, 호흡 곤란이 주요 증상이고, 두통, 오한, 인후통, 콧물, 근육통, 식욕 부진, 구토, 복통, 설사 등의 증상이 수반될 수 있다. 메르스 감염의 합병증으로 호흡 부전, 패혈성 쇼크, 다발성 장기 부전 등이 나타나기도 한다. 때에 따라서 감염 초기에는 호흡기 증상이 나타나지 않을 수 있다. 프랑스에서 보고된 최초의 환자는 초기 증상으로 설사와 복통을 호소했고, 이후 호흡기 증상이 나타났다(Mailles et al., 2013; Assiri et al., 2013). 당뇨병, 만성 폐질환, 암, 신부전 등 기저 질환이 있거나 면역기능이 저하된 사람들은 메르스 코로나바이러스에 감염될 위험이 일반인에 비해 높고, 예후도 불량한 것으로 알려졌다(질병관리본부, 2015; ECDC, 2016). 동반감염과 낮은 알부민 농도는 중증감염의 예측변수로 알려졌다(Saad et al., 2014). 메르스 코로나바이러스의

잠복기는 최소 2일에서 최대 14일이며 평균 5일 정도이다. 치명률은 약 40%이며, 나이와 기저 질환 여부 등에 따라서 사망률이 증가한다(질병관리본부, 2015; ECDC, 2016). 우리나라의 경우 확진환자 186명의 잠복기가 평균 6.83일이었고, 나이가 많거나 호흡기 질환이 있을 때 사망률이 높아지는 것으로 나타났다(질병관리본부, KCDC, 2015).

메르스 감염은 특정 연령에 국한되어 있지 않으며 당뇨병, 신부전, 만성 폐 질환, 면역 결핍 질환 등 기저 질환을 앓는 사람에서 중증의 급성 호흡기 질환을 일으킨다. 증상은 1차 감염자가 2차 감염자보다 더 중한 것으로 나타났다. 모든 환자는 직·간접적으로 중동 지역과 관련이 있다. 지역사회에서의 전파에 대한 근거는 나타나지 않았으며, 가족 간 전파와 의료기관에서의 제한적 전파로 인한 유행이 보고되었다. 명확한 감염 경로는 밝혀지지 않았지만 사람 간 밀접접촉에 의한 전파가 보고되었으며(질병관리본부, 2015) 의료 서비스에 노출되는 것이 감염에서 가장 위험한 요소로 알려졌다(Petersen et al., 2014; Saad et al., 2014; Banik et al., 2015; Oboho et al., 2015). Chowell 등(2014)은 수학적 모형을 이용하여 병원 내 전파와 지역사회에서의 전파의 상대적인 기여도를 비교하여 병원 내 전파가 지역사회에서의 전파에 비해 4배 더 높은 수준으로 메르스 코로나바이러스 전파에 기여한다는 결과를 발표했고, 감염의 통제조치가 입원 환자에 집중되어야 함을 제안하였다(Chowell et al., 2014). 우리나라에서는 총 16개의 병원에서 186명의 확진환자가 발생하였다. 이들 중 44.1%는 병원에서 메르스 코로나바이러스에 노출된 환자들이고, 32.8%는 환자를 병간호한 사람, 13.4%는 의료인이었다. 환자 중 83.2%가 다섯 건의 주요 전파사건(Super-Spreading Event, SSE)과 역학적 관련성이 있었다(질병관리본부, KCDC, 2015).

2) 메르스 바이러스 환자발생 양상

2012년 9월 사우디아라비아에서 메르스 코로나바이러스 감염의 첫 사례가 정식으로 보고되었다. 2012년 4월부터 2016년 10월 31일까지 총 1,843명의 메르스 환자가 발생하였고, 이 중 705명이 사망하였다(메르스 포털). 메르스는 중증급성호흡기증후군(SARS)이 발생하여 전 세계적으로 8,000여 명이 사망한 이후 10년 만에 발생한 호흡기 감염병으로, 중증급성호흡기증후군(SARS)이 1년 만에 사라진 것에 비해 지속적으로 발생하고 있다(Banik et al., 2015).

대다수의 메르스 환자가 중동 지역에서 발생하였고, 특히 사우디아라비아와 아랍에미리트에서 많은 환자가 발생하였다. 중동 이외 지역에서 발생한 메르스 환자는 중동 지역을 직접 여행하였거나, 중동 지역을 여행한 환자와 직·간접적인 관련이 있다(ECDC, 2015).

메르스 바이러스 환자가 가장 많이 발생한 사우디아라비아의 감염자 수는 2016년 10월 29일까지 1,488명(전 세계 발생 수의 80.74%)이고 사망자는 622명(전 세계 사망자 수의 88.23%)이다. 환자 발생수가 다음으로 많은 한국에서는 감염자가 186명(전 세계 발생 수의 10.09%)이고, 사망자는 38명(전 세계 사망자 수의 5.39%)이다. 아랍에미리트에서는 88명의 환자가 메르스 감염자와 13명의 사망자가 발생하였다. 요르단은 36명의 환자와 14명의 사망자가 발생하였다. 이외 중동지역에서는 카타르, 오만, 이란, 쿠웨이트, 레바논, 예멘에서 메르스 환자가 발생하였고, 유럽지역은 영국과 프랑스에서 각 4명과 2명의 환자가 발생했으며, 각 3명과 1명의 사망자가 발생하였다(메르스 포털).

2. 텍스트 마이닝의 개념과 특성

1) 텍스트 마이닝 분석의 개념

본 연구에서 활용한 텍스트 마이닝(Text Mining)은 Feldman & Dagan(1995)에 의해 텍스트 데이터베이스 기반의 지식발견(Knowledge Discovery in Textual Database)이라는 개념으로 처음 언급되었으며 텍스트 기반 데이터베이스로부터 자연어 처리 기술을 바탕으로 기존에 알려지지 않은 유용한 패턴과 지식을 발견하기 위한 목적으로 언어(텍스트)와 기계적 알고리즘(분석)의 이해관계가 상충됨으로써 사용자가 관심을 가지는 정보를 자동적으로 추출하는 프로세스를 의미 한다(Hotho et al., 2005).

텍스트 마이닝 기술을 통하여 거대한 텍스트 문치에서 의미 있는 정보들만 추출해, 다른 정보들과의 연관성을 파악하며 각각의 텍스트들이 가진 범주를 찾아내는 등의 결과를 얻어낼 수 있다. 컴퓨터가 인간이 사용하는 언어(자연어)를 분석하고 그 안에 숨겨진 정보를 찾아내기 위해 대용량의 언어 소스와 통계적이고 규칙적으로 나열되는 알고리즘이 사용되고 있다.(조성우, 2011).

텍스트 마이닝은 데이터 마이닝과 유사한 개념이지만 데이터 마이닝이 관계형 데이터베이스나 XML(Extensible Markup Language)와 같은 구조화된 데이터들만을 처리할 수 있는 반면, 텍스트 마이닝은 텍스트 문서, 이메일, HTML(Hypertext Markup Language)와 같은 비구조적인 데이터를 처리할 수 있다는 점에서 차이가 있다.(Fan et al., 2006).

2) 텍스트 마이닝 분석의 특성

기존의 내용분석에서는 연구자들이 직접 텍스트를 분석하고 판단해내는 작업을 하였다. 이러한 활동은 정확한 결과를 낼 수는 있었겠지만, 그에 반해 연구자가 직접 코딩 작업을 하는 과정에서 연구자의 가치가 들어갈 위험성과 많은 분량의 데이터를 소화해내지 못하는 단점이 나타났다. 텍스트 마이닝 기법을 활용한다면, 이러한 단점을 극복하여 다양한 분류가 가능해지고 각 분류에 따른 다각적인 해석을 실시할 수 있다(감미아, 송민, 2012).

텍스트 마이닝의 연구 분야는 문서 요약, 문서 분류, 문서 군집, 특성추출 분야다. 데이터 마이닝의 관점에서 문서로부터 구조화된 정보를 추출하여 데이터베이스화 시키거나 일정한 규칙을 찾아내는 것이 가장 일반적으로 사용되고 있으며, 이를 바탕으로 사용자가 웹상에서 문서를 찾는 것을 도와주거나, 사용자 프로필을 만들거나 분석된 문서에 쓰인 자연 언어 식별, 대량 DB에서 문서의 분류 및 군집화, 문서 분류 정보를 이용한 문서 재해석, 논문과 보고서 요약, 문서번역, 시계열정보의 습득을 통한 시장 및 위험도 분석, 문서 색인, 문서추천, 대표적인 키워드나 토픽추출, 질의응답시스템, 대규모 문서탐색 등이 대표적으로 응용되는 분야라고 할 수 있다(안태성, 서형국, 이경일, 2004).

비 구조화된 형태의 문서들이 수집되면 전 처리 과정을 거쳐서 텍스트 분석이 가능한 형태로 변환된다. 전 처리 과정에서는 전체 텍스트를 형태소 등의 단위로 분해(Parsing)하고, 중복 텍스트를 처리하고, 오류를 교정하는 작업 등이 이루어지는데 분석결과의 신뢰성을 확보하기 위한중요한 단계이다. 전 처리를 거친 데이터는 정보의 추출, 요약, 범주화 등 다양한 방식으로 텍스트 분석이 가능하며, 텍스트 분석 결과는 시각화되거나 유용한 정보 또는 지식의 형태로 도출될 수 있다(김근형, 오성열, 2009).

3) 텍스트 마이닝의 내용 및 과정

텍스트 마이닝은 크게 자료처리과정(Data Processing)과 자료 분석(Data Analysis)으로 나눌 수 있다. 자료처리과정은 비구조화 데이터를 분석에 용이하도록 가공 및 정제하는 단계이고 자료 분석은 데이터 마이닝, 기계학습(Machine learning), 통계학(Statistics) 등을 활용하여 텍스트로부터 유의미한 정보를 추출하는 단계이다(Hotho et al., 2005).

텍스트 마이닝의 자료처리과정(Data Processing)은 정보 검색(Information Retrieval, IR), 정보추출(Information Extraction, IE), 자연어 처리(Natural Language Processing, NLP) 등의 절차를 통하여 수집한 데이터를 가공 및 정제하는 과정이다(Hotho et al., 2005). 정보 검색은 사용자가 원하는 키워드를 기반으로 원하는 정보가 포함된 텍스트 데이터가 들어있는 문서(document)를 탐색하는 것이다. 정보 검색은 사용 목적에 따라 웹 검색(web search), 개인 정보 검색(personal information retrieval), 기업이나 기관, 특정 영역 검색(enterprise, institutional, and domain-specific search) 등 세 가지의 형태로 구별된다(Manning et al., 2008). 정보 추출은 토큰화(tokenization), 문장 분할(sentence segmentation), 품사의 배치(part-of-speech assignment)와 독립적 개체들로서의 인식 (identification of named entities)을 포함하는 일련의 과정을 거치는 것으로, 정보 검색이 사용자가 필요한 정보와 관련된 텍스트가 들어있는 특정한 문서를 찾는 것인데 반해 정보 추출이란 특정한 문서로부터 구체적인 정보를 정제하는 것을 말한다. 텍스트 마이닝의 자연어 처리는 구조와 형태가 복잡한 자연어를 컴퓨터로 분석하기 위해 가공하는 단계이며 한국어나 영어와 같이 문맥에 의존하는 자연어를 대상으로 하기 때문에 자료 처리과정에서 자연어 처리는 필수적이다. 자연어 처리는 크게 ‘형태소 분석’, ‘동사 분석’, ‘의미 분석’, ‘화용분석’으로 나뉜다. ‘형태소 분석’은 하나의 문장을 분해 가능한 최소한의 단위로 분리하고 분석하는 것을 말하는데 자연

어 처리에서 형태소 분석은 어휘 사전(lexicon)을 기반으로 입력문자의 형태소를 분석(matching)하는 것이다(Hotho et al., 2005).

자료처리과정이 원시자료의 가공 단계라고 한다면 자료 분석은 데이터 마이닝, 기계학습, 통계학 등을 활용하여 의미 있는 결과를 도출하는 과정 이다(Jarman, 2011). 즉, 텍스트로부터 의미 있는 추세와 패턴 및 지식을 발견하기 위하여 데이터 마이닝, 기계학습, 통계학 등을 활용하는 과정을 말한다(Hotho et al., 2005).

텍스트 마이닝과 같이 비구조화 데이터를 다루는 경우에 자료 분석 못지않게 자료 처리과정도 중요하기 때문에 자료 처리 과정상에서 다양한 자료 분석의 기법들을 적용하여 자료 처리과정의 성능을 높이는 연구가 활발히 진행 중에 있다(정다미 등, 2014; 최성필 등, 2009). 최근에는 비구조화 데이터의 자료 처리와 분석을 용이하게 할 수 있도록 대부분의 주요 통계 패키지과 데이터마이닝 프로그램들이 텍스트 마이닝 기능을 포함하고 있으며 텍스트 마이닝과 관련한 전처리 (Preprocess), 연관분석(Associate), 군집분석(Cluster), 요약(Summarize), 분류(Categorize)와 같은 다양한 분석 기능까지 제공하고 있다(Davi et al., 2005; Feinerer et al., 2008). 전처리(Preprocess)는 데이터를 준비하고 데이터를 해당 소프트웨어에 삽입, 가공 및 정제하는 일련의 전처리 과정이다. 연관분석(Associate)은 단어들 간의 동시 발생하는 빈도수를 바탕으로 연관성을 찾아내는 것이며, 군집분석(Cluster)은 유사한 객체들끼리 군집하여 동일한 그룹을 묶는 것이다. 요약(Summarize)은 텍스트에서 중요한 개념들을 요약한 것으로 일반적으로 빈도수가 높은 단어들을 반환하는 것이며, 범주화(Categorize)는 사전 정의된 범주로 텍스트들을 분류하는 것을 의미한다(Feinerer et al., 2008).

3. 관련 연구

1) 정부의 위기 대응 분석 연구

메르스 사태와 관련된 정부의 대응을 위험 커뮤니케이션 관점에서 분석한 연구들이 진행되었다. 세계보건기구(WHO, World Health Organization)는 위험 커뮤니케이션을 일컬어 위험 평가자, 위험 관리자, 이해관계자들 사이의 위험에 대한 의견 교류 과정이라고 정의했다.(2011) 이승아(2016)는 질적 연구 분석을 통해 메르스 기간 동안 관리당국의 브리핑 전문을 대상으로 위험 커뮤니케이션을 분석하였다. 해외 사례를 검토하여 위험 커뮤니케이션의 전략과 틀을 구성하여 분석하고 시사점을 도출하였다. 김은성(2016)은 메르스 관련 정부 위험소통의 한계에 대한 사회적 원인에 대해 결핍모델(Deficit Model) 및 맥락적 모델(Contextual Model) 관점을 통해 분석하였다. 결핍 모델은 정확한 과학적 지식을 대중에게 전달하는 것으로 과학적 지식이 부족한 대중과 전문가를 엄격히 분리하며 대중은 과학적 계몽의 대상이다. 반면 맥락적 모델에서의 대중은 지식을 비판하고 성찰할 수 있는 존재로 해석한다. 김은성(2016)은 정부의 위험 커뮤니케이션이 결핍모델에 기반했다고 판단했다. 초기에 예상했던 것보다 감염력이 강하자 전문가들은 당황했고 과학적 불확실성이 높은 상황에서 국민들에게 정보를 제공하는 것을 늦췄다고 보았다. 병원명을 공개하는 것에 있어서는 보건복지부와 질병관리본부가 병원 및 전문가들과 이해관계를 가지고 있었던 것이 중요한 이유라고 추정했다. 정부에 대한 시민들의 신뢰가 낮았던 것은 결핍모델에 의한 소통을 더 어렵게 했는데, 정부가 안심하라고 할수록 시민들이 불안해했을 것이라고 보았다. 전형준(2016)은 김은성(2016)이 제안한 후속연구를 진행하였다. 실제로 위험 커뮤니케이션에 관여했던 정부 기

관의 커뮤니케이션 관계자, 정부의 메르스 정책 관계자, 전문가(주요 취재원), 지방자치단체의 메르스 대응 관계자 등을 정보원으로 선정해 인터뷰를 진행하여 메르스 위기에 대한 정부관계자, 지방자치단체 담당자와 전문가 등의 위험 커뮤니케이션 인식과 행동은 결핍모델과 맥락모델 중 어떤 것으로 해석될 수 있는지, 또한 메르스 위기 상황에서 정부의 커뮤니케이션이 원활하지 않았다는 지적이 있었음을 감안하여 메르스 위기로 인해 등장한 커뮤니케이션 환경은 평상시와 어떻게 달랐는지 연구하였다. 메르스 관련 위기 대응 메시지를 언어 네트워크 분석(Semantic network analysis)을 통한 연구도 진행되었다. 이미나, 홍주연(2016)은 정부의 위기 대응 메시지를 언어 네트워크를 통해 분석하였다. 분석 시기를 3단계로 구분하여 중앙정부 및 지방정부의 보도자료를 대상으로 어절분석 및 네트워크 분석을 수행하였다.

2) 언론 보도 관련 연구

한편, 메르스 사태를 언론 보도의 프레임 개념을 이용하여 분석한 연구들이 진행되었다. 프레임 개념은 언론의 현실묘사 메커니즘을 밝히는데 유용한 분석틀로 자리 잡고 국내에 서도 활발한 연구가 진행되고 있다. 구체적으로 뉴스 프레임 연구는 언론이 현실을 있는 그대로가 아닌 특정한 관점이나 방식인 프레임을 통해 전달한다는 것을 전제로 하고 뉴스 텍스트로부터 프레임 유형을 밝히고 있다(이희영, 2016). 이희영(2016)은 지금까지 축적된 국내 뉴스 프레임 연구를 대상으로 메타 분석을 통해 뉴스 프레임 유형을 개발하였다. 나아가 메르스 사태 관련 방송뉴스를 대상으로 앞서 도출한 국내 언론의 뉴스 프레임 유형을 적용해 프레임을 분석하고 그 외연을 확장하였다. 지승재(2016)는 메르스 기간 동안 조선일보와 한겨레에 실린 메르스 관련 사설 및 칼럼을

수집하여 보수신문과 진보신문의 보도 프레임의 차이를 분석하였다. 조선일보의 경우 메르스라는 위기 상황을 해결하거나 대처하는 방법에 대해 다룬 반면 한겨레는 메르스라는 중심 이슈와 사회 구조적 맥락을 결부시키는 방법을 이용하였다. 안은영(2016) 조선일보와 한겨레의 메르스 관련 모든 기사를 추출하여(조선일보 571건, 한겨레 468건) 총 1,039건의 기사를 대상으로 뉴스 프레임의 양적인 내용분석(Contents analysis)을 실시하여 분석하였다. 내용분석은 말이나 글로 표현된 인간의 의사소통이 말하는 사람과 듣는 사람간 문화적 배경, 맥락, 시간, 언어 등의 요소로 왜곡될 수 있다는 가능성을 전제로 애매한 내용에 대한 인상적인 판단보다는 수적 용어를 사용하여 보다 간명하게 내용을 구분하는 방법이다(Krippendorff, 1980, Weber, 1985), 조선일보는 정부의 대처 방안 및 특정 사례 소개에 집중한 반면 한겨레는 메르스로 인한 혼란 및 피해의 책임 주체(주로 정부)를 강력하게 비판하면서 위험 상황을 전달하는데 초점을 맞추었다. 양현주(2016)는 메르스 관련 국내 일간지 보도 자료(조선일보, 동아일보, 한겨레, 경향신문) 2,385건을 기간별로 분류하여 내용 분석 방법을 사용하여 분석하고 보도 형태와 개선방안을 모색하였다. 차오잔원(2016)은 한국과 중국의 대표 일간지인 조선일보와 환구시보를 대상으로 각각 200건의 메르스 관련 기사를 연역적 접근방법에 따라 내용적인 측면에서 프레임 분석을 수행하였다. 뉴스 프레임 분석 외에 언론 보도에 대한 의미연결망 분석도 수행하였다. 권호천(2016)은 메르스 관련 조선일보와 한겨레신문 기사를 3단계 시기별로 나누어 의미연결망 분석을 수행하였다. 조선일보는 메르스가 확산되어 국내에서 발생하는 사망자와 감염자들에 대한 내용을 중심으로 이슈를 만들어 가고 있는 반면 한겨레신문은 기본적인 정보전달과 더불어 책임소재를 중심을 두고 이슈를 만들어 간 것으로 안은영(2016)의 연구결과와 비슷한 양상을 나타내었다.

3) 공중의 반응 관련 연구

메르스 사태에 대한 공중의 반응과 관련된 연구도 진행되었다. 박단비(2016)는 모바일 기반의 SNS(Social Networks Service)인 인스타그램(Instagram)에 게시된 ‘#메르스’ 해시태그와 관련된 이미지와 게시글을 수집하여 시기별 텍스트, 이미지를 분석하고 언어 프레임과 이미지 프레임의 구성을 도출하였다. 원태홍 등(2015)은 트위터, 페이스북과 같은 SNS상에서 메르스 첫 확진환자가 발표된 5월 20일부터 7월 31일까지 메르스 관련 SNS 멘션 중 이슈어를 선정하여 메르스 환자수의 변화와 관련 이슈어의 변화 양상을 비교 분석하였다. 헤럴드경제와 서울대학교 빅데이터 연구원 데이터 저널리즘 센터는 메르스 환자의 첫 확진일인 5월 20일부터 7월 14일까지 7주(총 56일)에 걸쳐 포털사이트 네이버를 통해 송출된 보수 및 진보 성향 중앙일간지 5개 매체의 기사 1만 1467건과 해당 기사에 달린 댓글 21만3901건을 빅데이터로 분석했다.

4) 총체적 연구의 필요성

이와 같이 메르스 사태와 관련된 정부의 위기대응 연구, 언론 보도에 대한 연구, 공중의 반응에 대한 선행 연구가 개별적으로 이루어졌다. 선행 연구에서는 공통적으로 정부의 위기 대응 및 소통이 문제점으로 지적되었다. 그러나 구체적으로 언론과 공중이 어떠한 대응과 소통을 원했는지 연구한 결과는 없었다. 이мина, 홍주연(2016)은 정부의 보도 자료가 언론에서는 어떤 내용으로 보도되었는지, 공중은 이에 대해 어떻게 반응하였는지 총체적으로 연구할 필요성을 제안하였다. 이에 본 연구는 메르스 관련 정부 보도자료 및 그에 따른 언론 및 공중 반응(뉴스 댓글)을 텍스트 마이닝을 활용하여 분석하고자 했다.

Ⅲ. 연구방법

1. 연구방법의 설정

선행 연구에서 메르스 관련 정부 보도 자료가 언론에서 어떤 내용으로 보도되었는지, 공중은 이에 대해 어떻게 반응하였는지 제안한 사항을 참고하여 연구 방법을 선정하였다(이미나, 홍주연, 2016).

연구방법 1 : 메르스 관련 정부 보도자료에 따른 언론 및 공중의 대응(뉴스기사 댓글)을 텍스트 마이닝으로 분석하여 주요 시기별로 빈번하게 언급하는 주요 단어들을 추출하여 단어 빈도(Term Frequency)를 분석하고자 한다. 이를 워드 클라우드(WordCloud)로 표현하여 메르스 사태 동안 정부, 언론, 공중의 관점이 각각 어떠한 양상을 나타냈는지 분석한다.

연구방법 2 : 메르스 관련 정부 보도자료에 따라 언론 및 공중이 주요하게 인식하는 단어들의 유사도(Similarity)를 파악하고자 한다. 보도자료, 언론기사, 댓글을 텍스트 마이닝으로 분석하여 추출된 각 단어의 코사인 유사도(Cosine Similarity)를 분석하고 주요 시기별로 비교해 보고자 한다. 이를 통해 메르스 사태 동안 정부, 언론, 공중의 관점이 유사했는지 여부를 통계적으로 유의한지 검증한다.

2. 연구 모형

본 연구는 <그림 1>과 같이 데이터 수집, 데이터 전처리, 데이터 분석 순으로 진행하였다. 우선 정부 보도 자료는 정부 메르스 공식 홈페이지(<http://www.mers.go.kr>)에 게시된 정부 보도 자료 총 286건 수집하였다. 중앙 정부에서 배포한 보도 자료는 전수 조사 하였으나 지방 정부에서 배포한 보도 자료는 연구 대상에서 제외하였다. 정부 보도에 따른 언론의 반응은 메르스 기간 동안 국내 최대 포털 사이트인 네이버(<http://www.naver.com>)에 게시된 메르스 관련 뉴스 349,217건 중 국가기간 언론사인 연합뉴스(<http://www.yonhapnews.co.kr>)의 기사 5,637건(1.61%)을 대상으로 수집하였다. 이 중 공중이 가장 관심 있게 구독했다고 추정되는 일일 하루 최다 댓글이 달린 1개 기사를 총 137건(2.43%)을 수집하였다. 공중의 반응의 경우 위 언론의 반응에 대한 댓글 394,914건 중 공감을 가장 많이 받은 411건(0.10%)의 댓글을 수집하였다. 검색 키워드는 ‘메르스’를 사용했으며 모든 자료는 메르스 최초 확진 환자가 발생한 2015년 5월 20일부터 메르스 유행이 종식된 2015년 12월 23일까지의 수집하였다. 시기별로 달라지는 양상을 보기위해 정부에서 설정한 초기 대응단계(2015년 5월 20일~6월 8일), 적극 대응단계(2015년 6월 9일~7월 27일), 후기 대응 및 복구단계(2015년 7월 28일~12월 23일)로 나누어 정부, 언론, 공중의 반응 자료를 수집하였다.

분석에 사용한 프로그램은 R 버전 3.3.1을 사용하였다 한글 명사 추출을 위한 KoNLR 패키지, 데이터 시각화를 위한 wordcloud 패키지 및 RColorBrewer 패키지, 데이터 처리를 위한 plyr 패키지 및 data.table 패키지, 문자열 처리를 위한 stringr 패키지를 사용하였다. 데이터 수집 이후 분석을 위한 전처리를 수행하였다. KoNLR 패키지의 extractNoun() 함수를 이용하여 명사를 추출하였

고 useSejongDic() 함수를 이용하여 고유명사를 사전에 추가하였다. stringr 패키지를 사용하여 숫자 및 문자를 불용어 처리 하였으며 gsub 함수를 사용하여 표준어가 아닌 용어를 모두 불용어 처리 하였다. 영문을 포함한 중복되는 단어는 모두 동의어 처리한 후 최종 데이터 셋을 구축하였다. 데이터 분석은 단어 빈도(Term Frequency)와 코사인 유사도(Cosine Similarity)를 사용하였다.

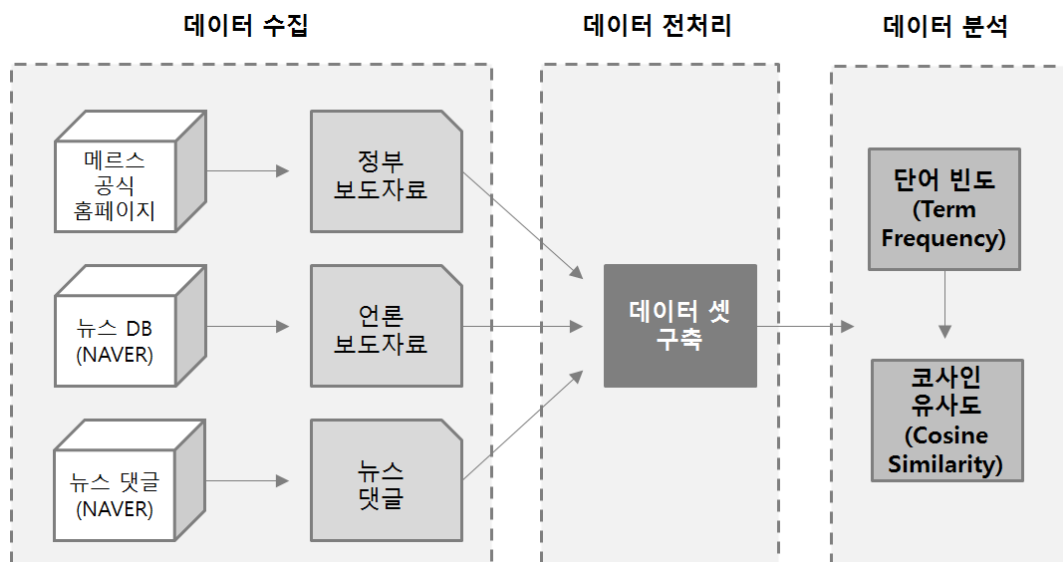


그림 2. 연구 모형.

3. 분석 방법

(1) 단어 빈도(Term Frequency)

본 연구에서 사용된 단어 빈도(Term Frequency)는 특정한 단어가 문서 내에 얼마나 자주 등장하는지를 나타내는 값으로, 이 값이 높을수록 문서에서 중요하다고 생각할 수 있다. 단어 빈도 $tf(t,d)$ 의 경우, 이 값을 산출하는 가장 간단한 방법은 단순히 문서 내에 나타나는 해당 단어의 총 빈도수를 사용하는 것이다. 문서 d 내에서 단어 t 의 총 빈도를 $f(t,d)$ 라 할 경우, 가장 단순한 tf 산출 방식은 (식 1) $tf(t,d) = f(t,d)$ 로 표현된다. 그 밖에 TF값을 산출하는 방식에는 다음과 같은 것들이 있다(C.D. Manning, Prabhakar and Hinrich, 2007).

- 불린 빈도: $tf(t,d) = 1$: t 가 d 에 한 번이라도 나타나면 1, 아니면 0;
- 로그 스케일 빈도: $tf(t,d) = \log(f(t,d) + 1)$;
- 증가 빈도: 문서의 길이에 따라 단어의 빈도값 조정

$$(식 2) \quad tf(t,d) = 0.5 + \frac{0.5 \times f(t,d)}{\max\{f(w,d) : w \in d\}}$$

(2) 코사인 유사도(Cosine Similarity)

코사인 유사도(Cosine similarity)는 내적공간의 두 벡터간 각도의 코사인값을 이용하여 측정된 벡터간의 유사한 정도를 의미한다. 각도가 0° 일 때의 코사인값은 1이며, 다른 모든 각도의 코사인값은 1보다 작다. 따라서 이 값은 벡터의 크기가 아닌 방향의 유사도를 판단하는 목적으로 사용되며, 두 벡터의 방향이 완전히 같을 경우 1, 90° 의 각을 이룰 경우 0, 180° 로 완전히 반대 방향인 경우 -1의 값을 갖는다. 이 때 벡터의 크기는 값에 아무런 영향을 미치지 않는다. 코사인 유사도는 특히 결과값이 $[0,1]$ 의 범위로 떨어지는 양수 공간에서 사용된다. 코사인 유사도는 어떤 개수의 차원에도 적용이 가능하여 흔히 다차원의 양수 공간에서의 유사도 측정에 자주 이용된다. 예를 들어 정보 검색 및 텍스트 마이닝 분야에서, 단어 하나 하나는 각각의 차원을 구성하고 문서는 각 단어가 문서에 나타나는 회수로 표현되는 벡터값을 가진다. 이러한 다차원 공간에서 코사인 유사도는 두 문서의 유사를 측정하는 매우 유용한 방법이다(Singhal, Amit, 2001). 코사인 유사도는 데이터 마이닝 분야에서 클러스터들간의 응집도를 측정하는 방법으로도 사용된다(P.-N. Tan, M. Steinbach & V. Kumar, 2005). 코사인 거리(cosine distance)라는 개념이 대신 사용되기도 하는데, 이는 (식 3) $D_C(A, B) = 1 - S_C(A, B)$ 으로 표현된다. 단, 코사인 거리는 삼각부등식의 성질을 갖고 있지 않으며 코사인 공리도 만족하지 않기 때문에 정확한 함수로 사용할 수는 없다. 순서를 유지하면서 삼각부등식 성질을 만족시키기 위해서는 각거리를 이용해야 한다. 코사인 유사도가 널리 사용되는 이유 중 하나는 이것이 양수 공간이라는 조건만 만족하면 얼마나 많은 차원 공간에서든지 거리를 측정하는 것이 가능하기 때문이다.

두 벡터의 코사인값은 유클리디안 스칼라곱 공식에서 유도할 수 있다.

$$(식 4) \mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta$$

속성 A, B의 벡터값이 각각 주어졌을 때, 코사인 유사도 $\cos(\theta)$ 는 벡터의 스칼라곱과 크기로 다음과 같이 표현할 수 있다.

$$(식 5) \quad \text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

이렇게 계산된 유사도는 -1에서 1까지의 값을 가지며, -1은 서로 완전히 반대되는 경우, 0은 서로 독립적인 경우, 1은 서로 완전히 같은 경우를 의미한다. 텍스트 매칭에 적용될 경우, A, B의 벡터로는 일반적으로 해당 문서에서의 단어 빈도가 사용된다. 코사인 유사도는 문서들간의 유사도를 비교할 때 문서의 길이를 정규화하는 방법의 하나라고 볼 수도 있다. 정보 검색의 경우, 문서의 단어 빈도가 음의 값이 되는 것이 불가능하므로 두 문서 사이의 코사인 유사도는 0에서 1까지의 값으로 표현되며, 두 단어 빈도간의 각도는 90° 를 넘을 수 없다.

IV. 연구결과

1. 분석자료의 일반적 특성

<표 1>은 메르스 관련 정부 보도 자료, 언론 보도 자료, 공중의 반응에 대한 일반적 특성이다. 모든 자료는 2015년 5월 20일부터 12월 23일까지 게시된 자료를 수집하였으며 정부에서 분류한 초기 대응단계(2015년 5월 20일 ~ 6월 8일), 적극 대응단계(2015년 6월 9일 ~ 7월 27일), 후기 대응 및 복구단계(2015년 7월 28일 ~ 12월 23일)로 나누어 수집하였다.

정부 보도 자료의 경우 중앙 정부의 보도자료 286건을 전수조사 하였으며 일반 보도자료, 보도자료 해명, 일일 현황보고로 분류하여 나타냈다. 정부의 메르스 보도 자료는 일일현황보고 95건(47.90%) > 일반보도자료(33.22%) > 보도해명자료(18.88%) 순으로 나타났다. 초기 대응단계에는 일반 보도자료(46.55%) > 일일현황보고(29.31%) > 보도 자료해명(24.14%) 순으로 나타났다. 메르스 사태 이후 초동 대응에 대한 일반 보도 자료가 다수였으며 일일 현황 보고는 메르스 유행 시작 31일 만인 2015년 6월 20일에 시작하였다, 각종 의혹에 대한 보도자료 해명 기사도 다수 보도하였다. 적극 대응단계에는 일일현황보고(39.13%) > 일반보도자료(34.06%) > 보도해명자료(26.81%) 순으로 나타났다. 48일간 54건의 일일현황보고를 보도하며 적극적인 대응을 보였다. 후기 대응 및 복구단계에는 일일현황보고(73.33%) > 일반보도자료(23.33%) > 보도해명자료(3.33%) 순으로 149일간 66건의 일일현황보고를 보도 하였다. 이 시기에는 보도해명자료가 급격히 줄어들었다. 시기별 정부 보도자료 건수는 적극 대응 단계 138건(48.25%) > 후기 대응 및 복구단계 90건(31.47%) > 초기 대응 단계 58건(20.28%) 순으로 적극 대응 단계에 가장 많은 보도 자료를 배포한 것으로

나타났다. <그림 2>를 통해 단계별 메르스 관련 정부 보도자료 건수의 증감을 확인할 수 있다. 일반 보도 자료와 보도자료 해명의 경우 적극 대응단계에 가장 많은 기사를 배포하였으며, 일일현황보고의 경우 후기 대응단계로 갈수록 증가하는 추세를 보였다.

표 1. 연구대상 자료의 일반적 특성

		초기 대응단계 (05.20-06.08)	적극 대응단계 (06.09-07.27)	후기 대응 및 복구단계 (07.28-12.23)	합 계
정부 보도자료	일반 보도자료	27 (46.55%)	47 (34.06%)	21 (23.33%)	95 (47.90%)
	보도 해명자료	14 (24.14%)	37 (26.81%)	3 (3.33%)	54 (33.22%)
	일일 현황보고	17 (29.31%)	54 (39.13%)	66 (73.33%)	137 (18.88%)
	소 계				286
언론 보도자료	보도자료 건수	1,721 (30.53%)	3,405 (60.40%)	542 (9.62%)	5,637
	댓글이 달린 보도자료 건수	752 (33.07%)	1,346 (59.19%)	181 (7.96%)	2,274
공중의 반응		184,920 (46.83%)	195,214 (49.43%)	14,780 (3.74%)	394,914

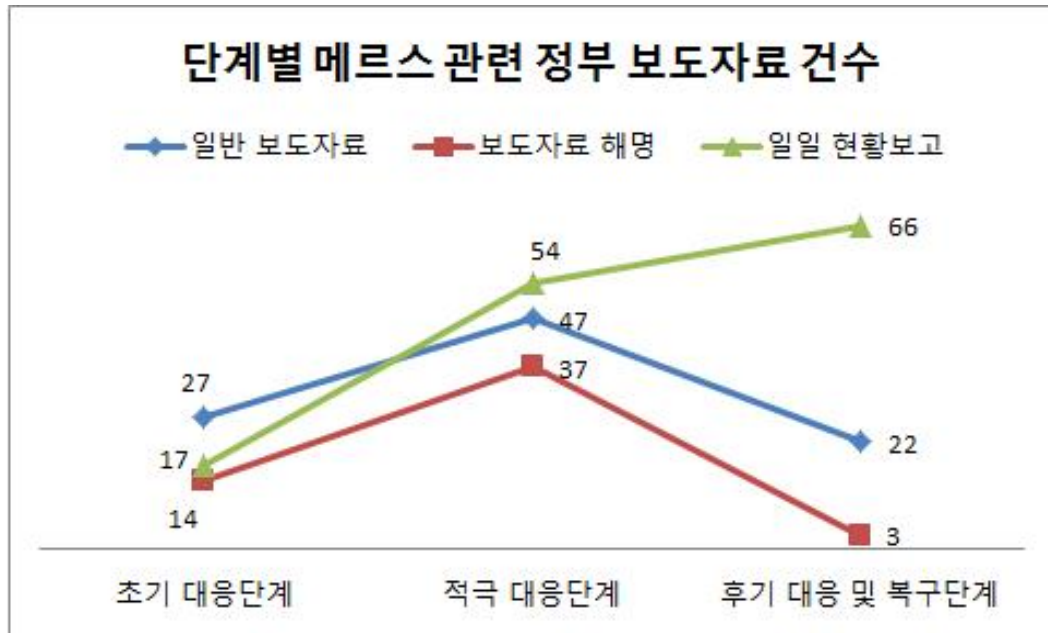


그림 2. 단계별 메르스 관련 정부 보도자료 건수.

언론 보도 자료의 경우 국가기간 뉴스통신사인 연합뉴스에서 게시한 보도 자료 5,637건 중 공중이 관심 있게 구독한 것으로 추정하는 댓글 달린 뉴스기사 수는 2,274건으로 나타났다. 단계별 기사 수는 적극 대응단계 3,405건 (60.40%) > 초기 대응단계 1,721건(30.53%) > 후기 대응 및 복구단계 542건 (9.62%) 순으로 적극 대응단계에 가장 많은 보도 자료를 배포한 것으로 나타났다. 댓글 달린 뉴스 기사 수 역시 적극 대응단계 1,346건(59.19%) > 초기 대응단계 752건(33.07%) > 후기 대응 및 복구단계 542건(9.62%) 순으로 나타났다. <그림 3>을 통해 단계별 메르스 관련 1개 언론기관의 보도자료 건수의 증감을 확인할 수 있다. 뉴스기사 및 댓글이 달린 뉴스기사 수 모두 적극 대응단계 > 초기 대응단계 > 후기 대응 및 복구단계 순으로 나타났다. 적극 대응단계에 메르스에 대한 관심이 가장 많은 것으로 볼 수 있다.

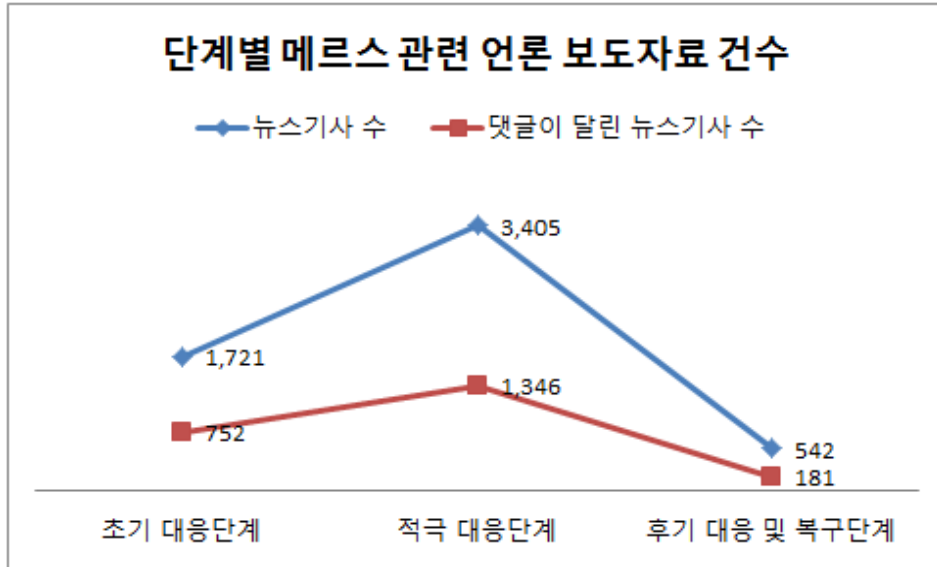


그림 3. 단계별 메르스 관련 언론 보도자료 건수.

공중의 반응의 경우 국가기간 뉴스통신사인 연합뉴스에서 게시한 보도 자료에 달린 전체 댓글 건수는 394,914건으로 나타났다. 전체 댓글 건수는 초기 대응단계 184,920건(46.83%)과 적극 대응단계 195,214건(49.43%)으로 비슷하게 나타났다. 기사 댓글의 경우 <그림 4>를 통해 초기 대응단계와 적극 대응단계의 건수가 비슷한 것을 확인 할 수 있다. 기사 건수가 적극 대응단계에 많았던 것을 비추어 보았을 때, 초기 대응단계에 공중의 관심이 많았다고 추측할 수 있다.

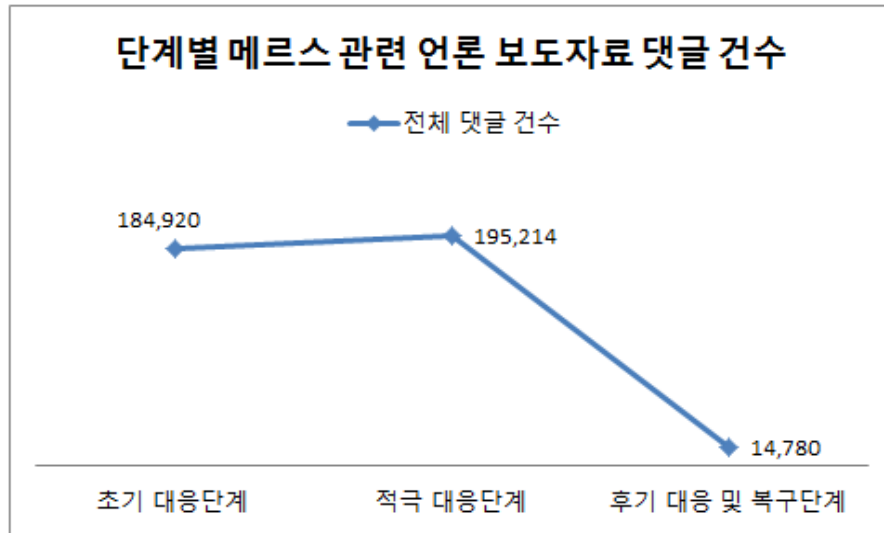


그림 4. 단계별 메르스 관련 언론 보도자료 댓글 건수.

본 연구에서 분석 대상으로 선정한 자료는 정부 보도 자료의 경우 중앙 정부의 보도자료 286건을 전수조사 하였고, 언론 보도 자료의 경우 국가간 뉴스 통신사인 연합뉴스에 게시된 보도 자료 중 공중이 가장 관심 있게 구독했다고 추정되는 일일 하루 최다 댓글이 달린 1개 기사를 총 137건(2.43%)을 수집하였다. 공중의 반응의 경우 위 언론의 반응에 대한 댓글 394,914건 중 일일 공감을 가장 많이 받은 댓글 3건씩 총 411건(0.10%)의 댓글을 수집하였다.

2. 텍스트 마이닝 결과

1) 단어 빈도(Term Frequency)

<표 2>은 데이터 전처리 과정 후 단계별, 그룹별 단어 출현 빈도를 나타낸 것이다. 정부 보도 자료는 3,236개의 텍스트, 언론 보도 자료는 3,261개의 텍스트, 공중의 반응은 1,014개의 텍스트가 검출되었다. 공중의 반응이 상대적으로 텍스트 양이 적은 것은 보도 자료와 비교하여 문서 형태의 차이에서 오는 절대적인 텍스트 수의 차이 때문으로 추측한다. 또한 비표준어, 비속어, 욕설, 기호, 이모티콘 등이 다수 사용되어 불용어 처리된 케이스가 많았다. 초기 대응단계의 정부 보도자료는 912개(52.08%)의 텍스트, 언론 보도자료는 667개(38.09%)의 텍스트, 공중의 반응은 172개(9.99%)의 텍스트가 검출되었다. 적극 대응단계의 정부 보도자료는 1,484개(48.66%)의 텍스트, 언론 보도자료는 1,229개(40.30%)의 텍스트, 공중의 반응은 337개(11.05%)의 텍스트가 검출되었다. 후기 대응단계의 정부 보도자료는 840개(31.00%), 언론 보도자료는 1,365개(50.37%), 공중은 505개(18.63%)의 텍스트가 검출되었다.

표 2. 단어 빈도(Term Frequency)

	정부 보도자료	언론 보도자료	공중의 반응	합 계
초기 대응단계	912 (52.08%)	667 (38.09%)	172 (9.99%)	1,751 (23.31%)
적극 대응단계	1,484 (48.66%)	1,229 (40.30%)	337 (11.05%)	3,050 (40.61%)
후기 대응단계	840 (31.00%)	1,365 (50.37%)	505 (18.63%)	2,710 (86.08%)
합계	3,236	3,261	1,014	7,511

(1) 초기 대응단계

<표 3>는 초기 대응단계의 정부 보도 자료, 언론 보도 자료, 공중의 반응에 대한 단어 빈도 분석 결과이다.

정부 보도 자료의 경우 가장 높은 빈도를 나타내는 단어는 메르스(308건) 이고, 그 다음으로 환자(263건) > 격리(141건) 등의 순으로 나타났다. 접촉, 병원, 감염, 검사, 의료기관, 확진, 증상, 보건복지부 등이 주로 언급되어 메르스 관련 병원이나 접촉자, 감염 여부, 확진 여부, 증상 등을 확인하는데 주목함을 알 수 있었다.

언론 보도자료의 경우 가장 높은 빈도를 나타내는 단어는 환자(137건) 이고, 그 다음으로 메르스(132건) > 감염(97건) 등의 순으로 나타났다. 격리, 접촉, 증상, 보건, 관리, 병원, 발생, 검사 등이 주로 언급되어 메르스 관련 병원 및 환자 등 감염 현황에 대한 내용에 주목하였다. 또한 대응, 관리, 대통령 등 정부의 대응에 대해서도 주목하였다.

공중의 반응의 경우 가장 높은 빈도를 나타내는 단어는 격리(8건) 이고, 그 다음으로 메르스(6건) > 박근혜(6건) 등의 순으로 나타났다. 대응, 무능, 감염 등 주로 정부의 대응에 대해 비판하는 것을 확인 할 수 있다.

표 3. 초기 대응단계 단어 빈도(Term Frequency)

순번	정부 보도자료		언론 보도자료		공중의 반응	
	단어	빈도	단어	빈도	단어	빈도
1	메르스	308	환자	137	격리	8
2	환자	263	메르스	132	메르스	6
3	격리	141	감염	97	박근혜	6
4	접촉	124	격리	92	대응	5
5	병원	122	접촉	54	무능	5
6	감염	116	증상	53	감염	4

7	검사	115	보건	49	건강보험	4
8	의료기관	113	관리	40	보건복지부	4
9	확진	108	병원	36	기사	4
10	증상	104	발생	32	문형표	4

(2) 적극 대응단계

<표 4>는 적극 대응단계의 정부 보도 자료, 언론 보도 자료, 공중의 반응에 대한 단어 빈도 분석 결과이다.

정부 보도 자료의 경우 가장 높은 빈도를 나타내는 단어는 확진(671건) 이고, 그 다음으로 메르스(526건) > 환자(523건) 등의 순으로 나타났다. 병원, 격리, 사망, 퇴원, 치료, 감염, 현황 등이 많이 언급되어 메르스 현황 파악에 주력하고 있음을 알 수 있다. 또한, 메르스 발생 병원이 공개되어 이에 대한 언급도 나타났음을 확인할 수 있다.

언론 보도 자료의 경우 가장 높은 빈도를 나타내는 단어는 환자(377건) 이고, 그 다음으로 메르스(326건) > 감염(153건) 등의 순으로 나타났다. 확진, 환자, 격리, 사망, 치료 등 환자 현황에 대한 내용에 주목하였으며, 삼성서울병원, 강동경희대병원 등 구체적인 병원에 대한 언급도 확인하였다. 당국, 보건복지부, 대통령, 중앙메르스관리대책본부 등정부의 대응에 대해서도 주목한 것을 확인하였다.

공중의 반응의 경우 가장 높은 빈도를 나타내는 단어는 메르스(35건) 이고, 그 다음으로 병원(14건) > 정부(12건) 등의 순으로 나타났다. 병원, 환자, 사망, 퇴원 등 메르스 현황에 대한 관심을 확인할 수 있었고 정부, 대통령, 박근혜, 문형표, 당국 등 정부의 대응에 대해 주목한 것을 확인하였다. 메르스에 대한 관리가 안정적인 국면에 들어서면서 감사, 고생, 다행 등 희망적인 메시지가 나타나기 시작했다.

표 4. 적극 대응단계 단어 빈도(Term Frequency)

순번	정부 보도자료		언론 보도자료		공중의 반응	
	단어	빈도	단어	빈도	단어	빈도
1	확진	671	환자	377	메르스	35
2	메르스	526	메르스	326	병원	14
3	환자	523	감염	153	정부	12
4	병원	496	격리	122	종식	11
5	격리	436	확진	103	의료진	11
6	사망	341	병원	100	환자	9
7	퇴원	338	삼성서울병원	100	삼성서울병원	9
8	치료	295	사망	87	사망	9
9	감염	227	치료	71	국면	9
10	현황	215	판정	70	대통령	6

(3) 후기 대응 및 복구단계

<표 5>는 후기 대응 및 복구단계의 정부 보도 자료, 언론 보도 자료, 공중의 반응에 대한 단어 빈도 분석 결과이다.

정부 보도자료의 경우 가장 높은 빈도를 나타내는 단어는 치료(322건) 이고, 그 다음으로 확진(303건) > 변동(249건) 등의 순으로 나타났다. 이는 후기 대응 단계에 가장 많았던 메르스 일일현황보고에 따른 결과로 추정된다. 관리, 대응, 협력, 강화, 방역, 조치 대비 등의 단어 빈도가 높은 것을 통해 대책 마련에 집중하고 있는 것으로 볼 수 있다.

언론 보도 자료의 경우 가장 높은 빈도를 나타내는 단어는 메르스(576건) 이고, 그 다음으로 환자(451건) > 감염(172건) 등의 순으로 나타났다. 환자, 감염, 병원, 격리, 치료 등을 통해 메르스 현황 파악에 주목하는 것을 확인할 수 있었다. 보건복지부, 관리, 정부 등 정부의 대응에 대한 관심과 병원, 서울대병

원, 삼성서울병원, 의료기관 등 의료기관의 대응에 대해 관심을 확인하였다. 종식이라는 단어가 등장하여 메르스 사태가 종료되는 국면에 들어선 것을 확인할 수 있었다.

공중의 반응의 경우 가장 높은 빈도를 나타내는 단어는 메르스(72건) 이고, 그 다음으로 병원(31건) > 사람(24건) 등의 순으로 나타났다. 정리, 종료를 통해 메르스 사태 종식에 대한 반응을 확인할 수 있었으며 감사를 통해 희망적인 메시지를 확인할 수 있었다. 반면 의문, 고발, 사죄 등 부정적인 메시지도 확인할 수 있었다.

표 5. 후기 대응단계 단어 빈도(Term Frequency)

순번	정부 보도자료		언론 보도자료		공중의 반응	
	단어	빈도	단어	빈도	단어	빈도
1	치료	322	메르스	576	메르스	72
2	확진	303	환자	451	병원	31
3	변동	249	감염	172	사람	24
4	퇴원	230	병원	156	정리	19
5	메르스	209	보건복지부	124	환경	15
6	환자	198	격리	95	다짐	14
7	사망	130	치료	92	진행	13
8	보건복지부	115	관리	85	의료	12
9	음성	112	판정	78	국민	12
10	감염	104	서울대병원	77	검사	11

2) 워드클라우드(WordCloud)

본 연구는 메르스 관련 정부 보도자료 및 그에 따른 언론 및 공중 반응(뉴스 댓글)을 분석하려는 것으로 정부 보도자료, 언론 보도자료, 공중의 반응에 대해 빈번하게 언급하는 주요 단어들을 시각화하기 위해 워드클라우드(WordCloud) 사용하여 나타내었다.

(1) 초기 대응단계

① 정부 보도자료

초기 대응단계 정부 보도자료의 빈출 단어는 메르스, 환자, 검사, 접촉, 감염, 의료기관, 검사, 증상, 보건복지부 등임을 알 수 있다.

- 32 -

③ 공중의 반응

후기 대응단계 공중의 반응의 빈출 단어는 메르스, 병원, 정부, 사람, 국민, 다행, 의사, 고생 등임을 알 수 있다.



그림 13. 후기 대응단계 공중의 반응 워드클라우드(WordCloud).

3. 코사인 유사도(Cosine Similarity) 분석

1) 기술통계 분석

정부 보도자료, 언론 보도자료, 공중의 반응의 코사인 유사도를 분석하기 전에 단어 빈도를 제외한 유사도를 확인하기 위해 단계별 기술통계 분석을 수행하였다. <표 6>은 중복 단어 결과를 나타낸 것이다.

표 6 중복 단어 결과

		정부 보도자료	언론 보도자료	공중의 반응
초기 대응단계	정부 보도자료	-	384(384/912) 42.11%	68(68/912) 7.46%
	언론 보도자료	384(384/667) 57.57%	-	71(71/667) 10.64%
	공중의 반응	68(68/172) 39.53%	71(68/172) 39.53%	-
	정부 보도자료	-	612(612/1,484) 41.24%	80(80/1,484) 5.39%
	언론 보도자료	612(612/1,229) 49.80%	-	160(160/1,229) 13.02%
적극 대응단계	공중의 반응	80(80/337) 20.67%	160(160/337) 47.47%	-
	정부 보도자료	-	507(507/840) 60.36%	165(165/840) 19.64%
	언론 보도자료	507(507/1,365) 37.14%	-	241(241/1,365) 17.66%
	공중의 반응	165(165/505) 32.67%	241(241/505) 47.72%	-
	정부 보도자료	-	507(507/840) 60.36%	165(165/840) 19.64%
후기 대응단계	언론 보도자료	507(507/1,365) 37.14%	-	241(241/1,365) 17.66%
	공중의 반응	165(165/505) 32.67%	241(241/505) 47.72%	-
	정부 보도자료	-	507(507/840) 60.36%	165(165/840) 19.64%
	언론 보도자료	507(507/1,365) 37.14%	-	241(241/1,365) 17.66%
	공중의 반응	165(165/505) 32.67%	241(241/505) 47.72%	-

(1) 초기 대응단계

정부 보도자료와 언론 보도자료의 경우 정부 보도자료의 관점으로 912건 중 384건으로 42.11% 만큼 유사했고, 언론 보도자료의 관점으로 667건 중 384건으로 57.57% 만큼 유사했다. 언론 보도자료와 공중의 반응의 경우 언론 보도자료의 관점으로 667건 중 71건으로 10.64% 만큼 유사했고, 공중의 반응 관점으로 172건 중 71건으로 39.53% 만큼 유사했다. 정부 보도자료와 공중의 반응의 경우 정부 보도자료의 관점으로 912건 중 68건으로 7.46% 만큼 유사했고, 공중의 반응 관점으로 172건 중 68건으로 39.53% 만큼 유사했다. 정부 보도자료와 언론 보도자료, 공중의 반응이 모두 중복된 단어는 55개로 나타났다. 정부 보도자료와 언론 보도자료의 단어 중복율은 44.34%로 나타났고(384/866), 언론 보도자료와 공중의 반응의 단어 중복율은 10.19%로 나타났으며(71/697), 정부 보도자료와 공중의 반응의 단어 중복율은 6.78%로 나타났다. (68/1,003)

(2) 적극 대응단계

정부 보도자료와 언론 보도자료의 경우 정부 보도자료의 관점으로 1,484건 중 612건으로 49.80% 만큼 유사했고, 언론 보도자료의 관점으로 1,229건 중 612건으로 49.80% 만큼 유사했다. 언론 보도자료와 공중의 반응의 경우 언론 보도자료의 관점으로 1,229건 중 160건으로 13.02% 만큼 유사했고, 공중의 반응 관점으로 337건 중 160건으로 47.47% 만큼 유사했다. 정부 보도자료와 공중의 반응의 경우 정부 보도자료의 관점으로 1,484건 중 80건으로 5.39% 만큼 유사했고, 공중의 반응 관점으로 337건 중 80건으로 20.67% 만큼 유사했다. 정부 보도자료와 언론 보도자료, 공중의 반응이 모두 중복된 단어는 64개로 나타났다. 정부 보도자료와 언론 보도자료의 단어 중복율은 50.58%로 나타났고(612/1,210), 언론 보도자료와 공중의 반응의 단어 중복율은 12.93%로 나타났으며(160/1,237), 정부 보도자료와 공중의 반응의 단어 중복율은 5.75%로 나타

났다(80/1,391).

(3) 후기 대응 및 복구단계

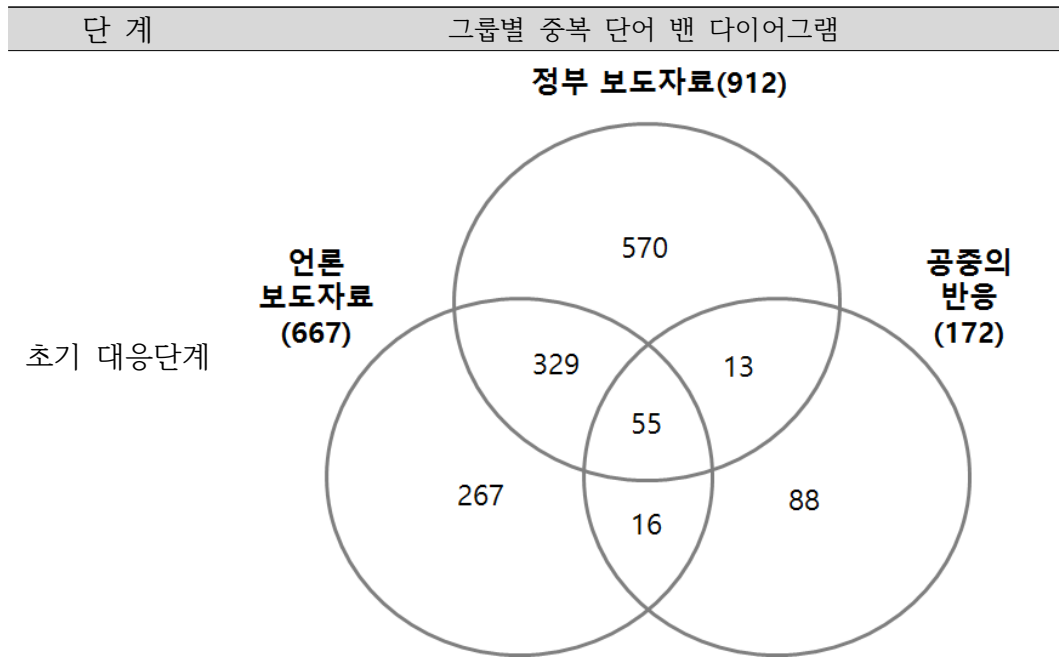
정부 보도자료와 언론 보도자료의 경우 정부 보도자료의 관점으로 840건 중 507건으로 60.36% 만큼 유사했고, 언론 보도자료의 관점으로 1,365건 중 507건으로 37.14% 만큼 유사했다. 언론 보도자료와 공중의 반응의 경우 언론 보도자료의 관점으로 1,365건 중 241건으로 17.66% 만큼 유사했고, 공중의 반응 관점으로 505건 중 241건으로 47.72% 만큼 유사했다. 정부 보도자료와 공중의 반응의 경우 정부 보도자료의 관점으로 840건 중 165건으로 19.64% 만큼 유사했고, 공중의 반응 관점으로 505건 중 165건으로 32.67% 만큼 유사했다. 정부 보도자료와 언론 보도자료, 공중의 반응이 모두 중복된 단어는 145개로 나타났다. 정부 보도자료와 언론 보도자료의 단어 중복율은 42.57%로 나타났고(507/1,191), 언론 보도자료와 공중의 반응의 단어 중복율은 17.36%로 나타났으며(241/1,388), 정부 보도자료와 공중의 반응의 단어 중복율은 16.26%로 나타났다. (165/1,015)

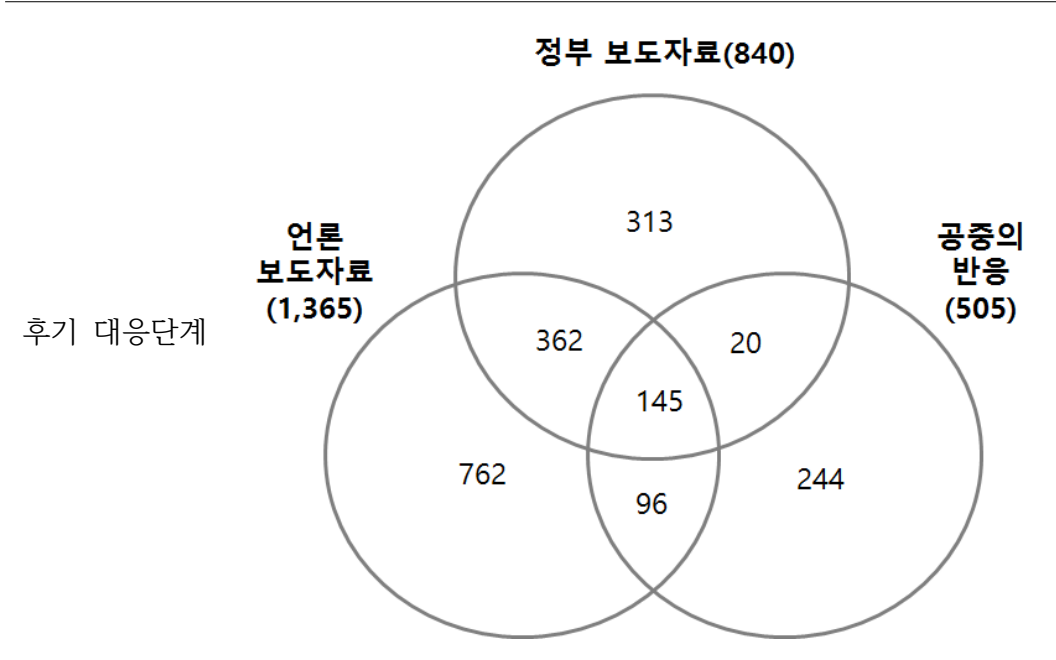
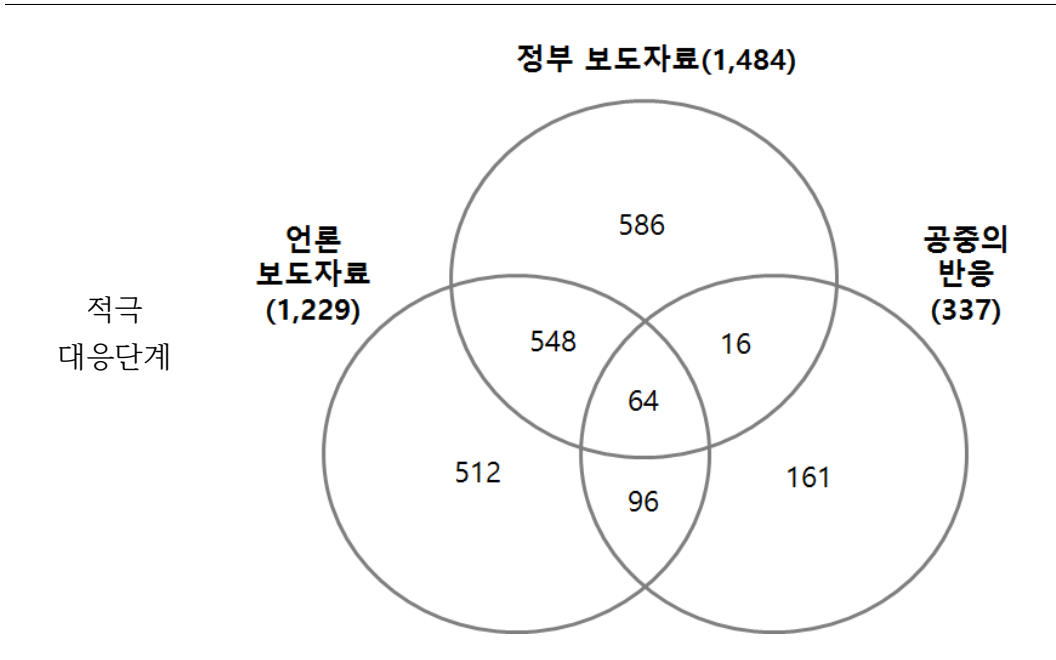
(4) 중복 단어 분석 결과

빈도를 제외한 중복 단어 분석 결과 초기 대응단계에서 정부와 언론의 메시지는 44.34%, 언론과 공중의 메시지는 10.19%, 공중과 정부의 메시지는 6.78%로 나타났다. 적극 대응단계에서 정부와 언론의 메시지는 50.58%, 언론과 공중의 메시지는 12.93%, 공중과 정부의 메시지는 5.75%로 나타났다. 후기 대응 및 복구단계에서 정부와 언론의 메시지는 42.57%, 언론과 공중의 메시지는 17.36%, 공중과 정부의 메시지는 16.26%로 나타났다. 초기, 중기, 후기 모두 정부-언론의 단어 중복율이 가장 높고 정부-공중의 단어 중복율이 가장 낮은 것으로 나타났다.

<표 7>은 단계별, 그룹별 중복 단어 벤 다이어그램을 나타낸 것이다. 초기 대응단계에서 정부와 언론에서 사용하지 않고 공중에서만 사용한 단어는 무능, 소 잃고, 외양간, 전염, 행동, 치료약 등으로 나타났다. 적극 대응단계에서 정부와 언론에서 사용하지 않고 공중에서만 사용한 단어는 세금, 문형표, 국회의원, 불신, 사퇴 등으로 나타났다. 후가 대응단계에서 정부와 언론에서 사용하지 않고 공중에서만 사용한 단어는 매뉴얼, 감기, 사죄, 용서, 명복 등으로 나타났다.

표 7 단계별, 그룹별 중복 단어 벤 다이어그램





2) 코사인 유사도 분석

정부 보도자료, 언론 보도자료, 공중의 반응을 단계별로 단어 빈도를 고려하여 코사인 유사도 분석을 수행하였다. 두 벡터의 코사인 값은 0~1을 가지며 벡터의 각이 작을수록 1에 가까워지고 클수록 0에 가까워진다. 즉 코사인 유사도가 1에 가까울수록 문서의 유사도가 높다고 볼 수 있다. <표 8>은 코사인 유사도를 나타낸 것이다.

(1) 초기 대응단계

정부 보도자료와 언론 보도자료의 유사도가 0.878로 가장 높았고 언론 보도자료와 공중의 반응의 유사도는 0.421, 정부 보도자료와 공중의 반응의 유사도는 0.389 순으로 나타났다.

(2) 적극 대응단계

정부 보도자료와 언론 보도자료의 유사도가 0.773으로 가장 높았고, 언론 보도자료와 공중의 반응의 유사도가 0.651, 정부 보도자료와 공중의 반응의 유사도는 0.531 순으로 나타났다

(3) 후기 대응단계

정부 보도자료와 언론 보도자료의 유사도가 0.599로 가장 높았고, 언론 보도자료와 공중의 반응의 유사도는 0.0977, 정부 보도자료와 공중의 반응의 유사도는 0.0843 순으로 나타났다.

표 8 코사인 유사도 분석 결과

		정부 보도자료	언론 보도자료	공중의 반응
초기 대응단계	정부 보도자료	-	0.878	0.389
	언론 보도자료	-	-	0.421
	공중의 반응	-	-	-
적극 대응단계	정부 보도자료	-	0.773	0.531
	언론 보도자료	-	-	0.651
	공중의 반응	-	-	-
후기 대응단계	정부 보도자료	-	0.599	0.0843
	언론 보도자료	-	-	0.0977
	공중의 반응	-	-	-

(4) 코사인 유사도 분석결과

초기 대응단계, 적극 대응단계, 후기 대응단계에 각각의 정부 보도자료와 언론 보도자료, 언론 보도자료와 공중의 반응, 정부 보도자료와 공중의 반응을 코사인 유사도를 이용하여 분석하였다. 분석 결과 모든 단계에서 정부 보도자료와 언론 보도자료 > 언론 보도자료와 공중의 반응 > 정부 보도자료와 공중의 반응 순으로 코사인 유사도가 높게 관찰되었다. 언론 보도는 비교적 정부 보도자료와 관심사가 비슷했으며, 공중의 반응은 비교적 정부 보도자료와 관심사가 다르다는 것을 통계적인 방법으로 확인하였다.

V. 고찰 및 결론

본 연구는 메르스 사태와 관련하여 정부 보도자료에 따른 언론 및 공중의 반응에 대해 텍스트 마이닝을 이용하여 분석하고 단계별, 그룹별로 메르스에 대한 관심사 얼마나 유사했는지 알아보기 위한 목적으로 수행하였다. 다양한 연구를 통해 메르스 사태를 통해 정부의 위기 대응 및 소통에 대한 문제를 제기하고 있지만 정부의 위기 대응에 대해 언론과 공중의 반응은 어떠한지, 즉 어떠한 소통을 원했는지 연구한 사례는 없었다.

연구 자료는 중앙정부 공식 홈페이지에 게재된 메르스 관련 보도자료 총 286건, 언론 보도자료 137건, 공중의 반응(언론기사 댓글) 411건을 초기 대응 단계(2015년 5월 20일~6월 8일), 적극 대응단계(2016년 6월 9일~7월 27일), 후기 대응 및 복구단계(2015년 7월 28일~12월 23일)로 시기를 나누어 수집하였다. 연구 결과는 다음과 같다.

첫째, 텍스트마이닝을 통한 단어빈도 및 워드클라우드 분석 결과 초기 대응 단계에서 정부는 주로 메르스 현황 파악에 집중하였고, 언론은 메르스 현황 및 정부의 대응에 주목하였으며, 공중은 주로 정부의 대응에 비판하는 경향을 보였다. 적극 대응단계에서 정부는 메르스 현황 파악 및 병원명 공개 등 적극적인 대응을 하였으며, 언론은 메르스 환자 현황 및 정부에 대응에 주목하였고, 공중은 메르스 현황, 정부의 대응, 희망적 메시지 등 다양한 의견을 나타냈다. 후기 대응 및 복구단계에서 정부는 지속적인 메르스 현황 보고 및 대책 마련, 복구 대응에 집중하였고, 언론은 메르스 현황 및 정부의 대응, 종식 등에 대해 주목하였으며, 공중은 메르스 종식에 따른 반성, 희망적 메시지, 부정적 메시지 등이 나타났다.

둘째, 빈도를 제외한 단어 중복을 분석 결과 초기 대응단계에서 정부와 언

론의 메시지는 44.34%, 언론과 공중의 메시지는 10.19%, 공중과 정부의 메시지는 6.78%로 나타났다. 적극 대응단계에서 정부와 언론의 메시지는 50.58%, 언론과 공중의 메시지는 12.93%, 공중과 정부의 메시지는 5.75%로 나타났다. 후기 대응 및 복구단계에서 정부와 언론의 메시지는 42.57%, 언론과 공중의 메시지는 17.36%, 공중과 정부의 메시지는 16.26%로 나타났다. 모든 단계에서 공중-정부, 언론-공중, 정부-언론 순으로 높게 나타났음을 확인하였다. 정부와 언론의 단어 중복율이 가장 낮았으며 이는 정부가 언론이 원하는 정보를 적절하게 제공하지 않았다고 볼 수도 있다.. 초기 대응단계에서 정부와 언론에서 사용하지 않고 공중에서만 사용한 단어는 무능, 소 잃고, 외양간, 전염, 행동, 치료약 등으로 나타났다. 적극 대응단계에서 정부와 언론에서 사용하지 않고 공중에서만 사용한 단어는 세금, 문형표, 국회의원, 불신, 사퇴 등으로 나타났다. 후기 대응단계에서 정부와 언론에서 사용하지 않고 공중에서만 사용한 단어는 매뉴얼, 감기, 사죄, 용서, 명복 등으로 나타났다. 공중의 반응은 정부 대응에 대한 비판적인 내용과 더불어 실제적인 행동지침과 매뉴얼, 건강보험에 대한 내용 등으로 나타났다.

셋째, 코사인 유사도 분석 결과 초기 대응단계, 적극 대응단계, 후기 대응 및 복구단계 모두 정부 보도자료와 언론 보도자료가 가장 높은 코사인 유사도를 나타냈으며, 정부 보도자료와 공중의 반응의 유사도가 가장 낮은 유사도를 나타냈다. 즉, 정부가 전달한 메시지와 공중이 원하는 메시지의 유사도가 비교적 낮았다.

선행 연구에서는 언어 네트워크 분석, 의미 연결망 분석 등을 사용하여 메르스 사태와 관련된 정부, 언론, 공중의 반응을 각각 연구하였다. 주요 시기별로 정부나 언론에서 어떠한 단어를 가장 많이 사용하였는지, 각 단어의 연관성을 얼마나 높은지 등을 분석하였다. 선행 연구와 본 연구를 비교한 결과는 다음과 같다.

첫째, 정부의 위기 대응 메시지를 언어 네트워크로 분석한 이미나, 홍주연(2016)의 연구에서는 분석 시기를 메르스 초기 단계(5월 20일 ~ 6월 1일), 확산기(6월 2일 ~ 6월 13일), 쇠퇴기(6월 14일 ~ 7월 6일)로 구분하여 초기 11건, 확산기 97건, 쇠퇴기 26건의 중앙정부 및 지방정부의 보도자료를 어절분석 및 네트워크 분석을 수행하였다. 본 연구결과와 비교한 결과 자료 수집 기준과 분석 시기가 상이하여 단어 빈도에 차이가 있으나 최빈도 단어는 대부분 일치하는 것으로 확인하였다.

둘째, 메르스 관련 언론 보도를 의미연결망 분석을 수행한 권호천(2016)의 연구에서는 분석 시기를 메르스 초기 단계(5월 20일~6월 3일), 확산단계(6월 4일~7월 11일), 진정단계(7월 12일~12월 23일)의 3단계 시기별로 나누어 조선일보와 한겨레신문 기사 제목을 분석하였다. 본 연구 결과의 단어 출현빈도를 비교한 눈에 띄는 차이점은 없었다. 다만 권호천(2016)의 연구의 경우 이념적 성향이 다른 두 언론사의 메르스 관련 보도자료를 의미연결망을 통해 분석했기 때문에 본 연구 결과와 직접 비교가 어려웠다.

셋째, 메르스 관련 공중의 반응을 분석한 헤럴드경제와 서울대학교 빅데이터 연구원 데이터 저널리즘 센터는 메르스 환자의 첫 확진일인 5월 20일부터 7월 14일까지 7주(총 56일)에 걸쳐 포털사이트 네이버를 통해 송출된 보수 및 진보 성향 중앙일간지 5개 매체의 기사 1만1467건과 해당 기사에 달린 댓글 21만3901건을 빅데이터로 분석했다. 본 연구결과와 비교한 결과 상위 빈도 단어 대부분 일치하는 것으로 확인하였다.

정부의 위기 대응 및 소통에 대한 분석을 위해서는 정부의 위기 대응에 대해 언론과 공중이 어떠한 반응을 보였는지 통합적으로 분석하는 것이 필요하다. 본 연구는 텍스트 마이닝을 통해 메르스 관련 정부의 대응, 즉 정부 보도 자료에 따른 언론 및 공중의 반응을 분석하였고 그 결과 정부의 대응과 공중의 반응은 상이하다고 볼 수 있다는 것을 도출했다는 것에 의의가 있다. 그러

나 본 연구는 다음과 같은 한계점이 있다.

첫째는 연구 자료의 대표성에 대한 한계점이다. 본 연구에서 수집한 정부 보도 자료의 경우 중앙 정부에서 배포한 보도 자료를 전수조사 하였다. 하지만 중앙 정부 자료 외 각 시와 지방정부에서 제공한 보도 자료를 누락하였다. 언론 보도 자료의 경우 국가기간 언론사인 연합뉴스의 하루 최대 댓글이 달린 1개 기사만을 수집하였다. 공중의 반응의 경우 댓글의 경우 최대 댓글이 달린 1개 기사의 베스트 댓글 3개만을 수집하여 공중의 반응을 대표하지 못한다는 점, 특히, 댓글은 주로 한정적인 집단이 작성하였다는 한계점이 있다. 향후에는 중앙 정부 외 지방 정부의 보도자료, 메르스 관련 주요 언론 보도자료, 더 많은 뉴스기사 댓글 및 SNS 메시지 등을 추가적으로 분석하여 메르스 사태에 대하여 중앙 정부-지방 정부의 대응에 대한 언론-공중의 반응을 총체적으로 분석할 필요가 있다.

둘째는 연구 방법에 대한 한계점이다. 본 연구는 코사인 유사도를 통해 정부 보도자료, 언론 보도자료, 공중의 반응의 유사도를 분석하였다. 문서 유사도를 분석하는 다른 방법으로 TF-IDF(Term Frequency-Inverse Document Frequency)가 있다. 이는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다. 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용할 수 있다. TF(단어 빈도, term frequency)는 특정한 단어가 문서 내에 얼마나 자주 등장하는지를 나타내는 값으로, 이 값이 높을수록 문서에서 중요하다고 생각할 수 있다. 하지만 단어 자체가 문서군 내에서 자주 사용되는 경우, 이것은 그 단어가 흔하게 등장한다는 것을 의미한다. 이것을 DF(문서 빈도, document frequency)라고 하며, 이 값의 역수를 IDF(역 문서 빈도, inverse document frequency)라고 한다. TF-IDF는 TF와 IDF를 곱

한 값이다(Salton and McGill 1983). 본 연구에서는 문서에 있는 텍스트를 추출하여 출현빈도와 함께 두 문서에 대해 코사인 유사도를 분석하였다. 향후 연구에서는 TF-IDF를 활용한 유사도 분석을 통해 코사인 유사도 분석 결과와 비교하여 유사도가 어떠한 양상으로 나타나는지 연구할 필요가 있다.

본 연구를 통해 메르스 관련 정부의 대응 메시지와 공중의 반응에 대한 유사도가 비교적 낮았음을 확인하였다. 향후에는 메르스 사태와 관련하여 정부의 대응과 언론 및 공중의 반응을 대표할 수 있는 자료를 수집하여 정부의 위기 대응 및 소통이 어떻게 이루어졌는지 더욱 객관적이고 정량화된 연구가 필요할 것이다.

참 고 문 헌

- 권호천. 메르스 사태에 대한 신문보도의 의미연결망 분석: 조선일보와 한겨레신문 기사를 중심으로. 의료커뮤니케이션 2016:11(1);63-80.
- 김근형, 오성열. 온라인 고객리뷰 분석을 통한 시장세분화에 텍스트마이닝 기술을 적용하기 위한 방법론. 한국콘텐츠학회 2009:9(8);272-84.
- 김은성. 메르스관련 정부 위험소통의 한계에 대한 사회적 원인 분석. 한국위기관리논문집 2015:11(10);91-109.
- 김미아, 송민. 텍스트 마이닝을 활용한 신문사에 따른 내용 및 논조 차이점 분석. 지능정보연구 2012:18(3);53-77.
- 네이버(URL: <http://www.naver.com>).
- 박단비. 메르스 질병 프레임 분석: 인스타그램 텍스트, 이미지 분석을 통하여. 석사학위 논문, 성균관대학교 대학원, 2015.
- 보건복지부. 감염병 위기관리 표준 매뉴얼. 2014.
- 보건복지부 · 질병관리본부. 2015 중동호흡기증후군(MERS) 대응지침(3판). 2015.
- 보건복지부 · 질병관리본부. 2015 메르스(MERS) 대응지침(3-3판). 2015.
- 보건복지부 · 질병관리본부. 2015 메르스(MERS) 대응지침(3-5판). 2015.
- 보건복지부 · 한국보건사회연구원 백서연구팀. 2015 메르스 백서. 2015.

연합뉴스(URL: <http://www.yonhapnews.co.kr>).

서울대학교 빅데이터 연구원 데이터 저널리즘
센터(URL:<http://bdi.snu.ac.kr/?p=3853>).

안은영. 중동호흡기증후군(MERS) 위험보도 프레임 연구 -조선일보와
한겨레신문을 중심으로. 석사학위 논문, 중앙대학교 대학원, 2016.
안태성, 서형국, 이경일. 텍스트마이닝 기반 고정밀 검색시스템.
정보처리학회지 2004:11(2);88-97.

양현주. 위험커뮤니케이션 사례분석 연구 -메르스 관련 보도 분석을 중심으로.
석사학위 논문, 호남대학교 대학원, 2016.

원태홍, 배규환, 유환희. 메르스 사태에 따른 SNS 이슈어 분석.
한국지형공간정보학회 학술대회 2015:9(1);244-245.

이미나, 홍주연. 메르스 확산에 따른 정부의 위기 대응 메시지 언어 네트워크
분석. 한국콘텐츠학회논문지 2016:5(1);124-136.

이승아. 정부의 위험 커뮤니케이션 전략에 관한 연구 -메르스 확산 사례를
중심으로. 석사학위 논문, 한양대학교, 2016.

이희영. 뉴스프레임 유형화 연구 -국내 연구 메타분석을 통한 유형 도출 및
메르스 방송뉴스에 대한 적용. 박사학위 논문, 한양대학교 대학원, 2016.

정다미, 김재석, 김기남, 허종욱, 온병원, 강미정. 사회문제 해결형 기술수요
발굴을 위한 키워드 추출 시스템 제안, 지능정보연구. 2013:19(3);1-23.

정부 메르스 공식 홈페이지(URL: <http://www.mers.go.kr>).

조성우. Big Data 시대의 기술. 2011.

중앙메르스관리대책본부. 메르스(MERS) 대응 통합 행정지침(1-2판). 2015.

중앙메르스관리대책본부. 메르스 접촉자 관리 매뉴얼. 2015.

지승재. 메르스 사태에 관한 사설과 칼럼의 프레임과 내용분석: 조선일보와 한겨레 비교 분석. 석사학위 논문, 건국대학교 대학원, 2016.

질병관리본부 중동호흡기증후군대책반. 중동호흡기증후군(MERS) 예방 및 관리지침(1판). 2014.

질병관리본부 중동호흡기증후군대책반. 중동호흡기증후군(MERS) 관리지침(2판). 2014.

차오잔원. 메르스(MERS) 이슈에 대한 한국과 중국 언론의 보도행태에 대한 프레임 연구 -조선일보와 환구시보를 중심으로. 석사학위 논문, 경기대학교 대학원, 2015.

최성필, 정창후, 최윤수, 맹성현. 평면적 어휘 자질들을 활용한 확장 혼합 커널 기반 관계 추출. 정보과학회논문지. 2009:16(11);1046-50.

Amit Singhal. Modern Information Retrieval: A Brief Overview. IEEE Data Engineering Bulletin 2001:24(4);35-43.

Assiri A, McGeer A, Perl TM, Price CS, Rabeeah AA, Cummings D, et al. Hospital Outbreak of Middle East Respiratory Syndrome Coronavirus. New England Journal of Medicine. 2013:369;407-416.

Banik GR, Khandaker G, Rashid H. Middle East Respiratory Syndrome Coronavirus “MERS-CoV” Current Knowledge Gaps. Paediatric Respiratory Reviews. 2015:16;197-202.

- Chowell G, Blumberg S, Simonsen L, Miller MA, Viboud C. Synthesizing data and models for the spread of MERS-CoV, 2013: Key role of index cases and hospital transmission. *Epidemics*. 2014;9;40-51.
- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. *Introduction to Information Retrieval*. 2007.
- Davi, A., Haughton, D., Nasr, N. Shah, G., Skaletsky, M. and Spack, R. A Review of Two Text-Mining Packages: SAS TextMining and WordStat, *The American Statistician*. 2005;59(1);89-103.
- ECDC(URL: <http://ecdc.europa.eu>)
- Feinerer I., Hornik K. and Meyer D. Text Mining Infrastructure in R, *Journal of Statistical Software*. 2008;25(5);1-54.
- Hotho, A., Nurnberger, A. and Paa, G. *A Brief Survey of Text Mining*. 2015.
- Jeon HJ, An Analysis of Risk Communication-A Case Study of MERS-CoV in Korea. *Crisisonomy* 2016;12(5);143-155.
- Jarman, J. *Combining Natural Language Processing and Statistical Text Mining: A Study of Specialized Versus Common- 121 Languages*. 2011.
- Korea Centers for Disease Control and Prevention(2015). Middle East Respiratory Syndrome Coronavirus Outbreak in the Republic of Korea. *Osong Public Health and Research Perspectives*. 2015;6(4);269-278.
- Klaus Krippendorff. *Validity in Content Analysis*. 1980.
- Mailles A, Blanckaert K, Chaud P, van der Werf S, Lina B, Caro V, et al. The investigation team. First cases of Middle East Respiratory Syndrome

Coronavirus (MERS-CoV) infections in France, investigations and implications for the prevention of human-to-human transmission, France. Eurosurveillance 2013;18(24).

Oboho IK, Tomczyk SM, Al-Asmari AM, Banjar AA, Al-Mugti H, Aloraini MS, et al. 2014 MERSCoV outbreak in Jeddah – a link to health care facilities. New England Journal of Medicine . 2015;372;846-54.

Petersen E, Pollack MM, Madoff LC. Health-care associate transmission of Middle East respiratory syndrome Corona virus, MERS-CoV, in the Kingdom of Saudi Arabia. International Journal of Infectious Diseases. 2014;29;299-300.

P. N. Tan, M. Steinbach, V. Kumar. Introduction to Data Mining. 2005.

Robert Philip Weber. Basic content analysis. 1985.

Sadd M, Omrani AS, Baig K, Bahloul A, Elzein F, Matin MA, et al. Clinical aspects and outcomes of 70 patients with Middle East respiratory syndrome coronavirus infection: a single-center experience in Saudi Arabia. International Journal of Infectious Diseases. 2014;29;301-6.

Salton, G; McGill, M. J. Introduction to modern information retrieval. 1986.

Weiguo Fan. Tapping the Power of Text Mining Communications of the ACM. 2016.

= ABSTRACT =

An Analysis of response of the press and public
according to government press releases related to
MERS using text mining

Jinki Jung

Department of Biostatistics

Graduate School of Public Health

Yonsei University

(Directed by Professor Jungmo Nam, Ph.D.)

Background

In recent years, the importance of the crisis response from the infectious diseases has been highlighted all over the world such as Severe Acute Respiratory (SARS), Ebola, Middle East Respiratory Syndrome (MERS), Zika virus, etc. The MERS crisis that occurred in May 2015 is regarded as a representative example of the failure of the government's crisis communication. The government has not adequately explained the path of the MERS risk to the public and was reluctant to disclose the name of the hospital where the infected person occurred. Citizens voluntarily began

exchanging risk information about MERS voluntarily through Social Network Service (SNS). The government then announced that it would punish those who spread inaccurate information about MERS. However, the announcement of this government has made the public angry.(Kim, 2015). This study analyzed the response messages of the government related to MERS to examine how the government perceived and responded to crises and how the responses of the press and the public were accordingly.

Subject and Methods

According to the official report of the government, the research data are divided into the initial response phase (May 20 - June 8, 2015), the active response phase (June 9 - July 27, 2016) and the late response and recovery phase (July 28 to December 23, 2015). A total of 286 press releases were collected on the official web-site of the central government. 137 press releases and public response (commented on press articles) 411 cases were collected. This was preprocessed and analyzed by text mining. It is analyzed. Through analysis of word frequency (Term Frequency), word cloud (Word Cloud), and cosine similarity analysis, government press releases, media press releases, and public reaction patterns and similarities were analyzed.

Results

As a result of term frequency and word cloud analysis by the time of the MERS crisis, the government focused on identifying the status of MERS, disclosing hospital names, preparing countermeasures, and responding to

recovery. The press paid attention to the status of the MERS, the response of the government, and the termination of MERS. The public criticized mainly the government's response, indicating the status of marts, government responses, hopeful messages, and reflection.

As a result of analysis of word overlapping rate except frequency, the overlap rate of government word and public word was lowest at all stages. It would be inferred that the government did not adequately provide the information the press and the public wanted.

As a result of analysis of cosine similarity, the degree of cosine similarity between government press releases and public responses was lowest at all stages. In other words, the similarity between the message delivered by the government and the message desired by the public was relatively low.

Conclusions

As a result of examining the press and public reactions according to the government press release according to the time of the spread of the crisis, the similarity of the response message of the government and the response of the public was relatively low. It would be seen as the government's inability to appropriately respond to the information the public wants. It would be viewed as the government's inability to properly provide the information that the public wants.

Although there is a limit to collecting research data, this study shows that there is a statistical difference between the response message of the

government and the response of the press and the public in a national disaster situation called MERS. It is meaningful that the result of providing the information properly is significant.